AD_____

AWARD NUMBER: DAMD17-03-1-0433

TITLE: Proteomic Analysis of Genistein Mammary Cancer Chemoprevention

PRINCIPAL INVESTIGATOR: Coral A. Lamartiniere

CONTRACTING ORGANIZATION: University of Alabama at Birmingham
Birmingham, Alabama 35294

REPORT DATE: July 2006

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-07-2006	Annual	1 Jul 2005 – 30 Jun 2006

4. TITLE AND SUBTITLE

Proteomic Analysis of Genistein Mammary Cancer Chemoprevention

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-03-1-0433

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Coral A. Lamartiniere

E-Mail: Coral@uab.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Alabama at Birmingham
Birmingham, Alabama 35294

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We have developed reproducible methodology for separation, biochemical and statistical analysis and identification of mammary gland proteins. We have discovered that GTP-CHI expression is significantly increased shortly following exposure at day 21 by genistein. At day 50, there was significant up-regulation of tyrosine hydroxylase and VEGF-R2. This and previous work suggests that early postnatal (prepubertal) exposure to genistein enhances cell proliferation and cell differentiation and gland maturation. This unique developmental maturation leads to a new biochemical "blue-print" whereby the cells have reduced EGF-signaling and VEGFR2 that render the mature mammary gland less proliferative and susceptible to chemically-induced mammary cancer initiation, angiogenesis and for cancer progression. Our experiments in identifying protein biomarkers in interstitial fluid surrounding mammary glands in rats have been mucli more diEcult and not successful to-date. We have recently developed our own probes that should allow us to collect a higher yield of proteins and hopefully allow success.

15. SUBJECT TERMS
Genistein, Breast Cancer Chemoprevention, Proteomics, Rats

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	23	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

INTRODUCTION

Breast cancer is the most common malignancy diagnosed in American women. Even with improved technology for early detection and aggressive therapeutics, most often the disease is incurable once it is discovered. We believe that prevention rather than therapy is the desired future against cancer, and that innovative approaches and new technology will be the key to breakthroughs. Towards this, our laboratory has been studying how developmental alterations to the mammary gland can program against this cancer. More specifically, we have demonstrated that prepubertal only, and prepubertal plus adult, exposure(s) to dietary genistein, a phytoestrogen component of soy, confer(s) a long-term protective effect against dimethylbenz(a)anthracene (DMBA)-induced mammary cancer in rats (1). Consistent with our findings are epidemiological reports that Asian women exposed to a diet high in soy during adolescence have a lower incidence of breast cancer (2, 3). We have hypothesized that genistein exerts its chemopreventive actions by postnatally programming developmental modifications to genes/proteins that render the mammary gland less susceptible to cancer. The objective of this proposed research is to identify regulatory proteins responsible for conferring breast cancer protection using innovative aims and technology.

BODY

Specific Aim 1) To identify proteins that are differentially expressed in mammary glands of rats treated ± the carcinogen, DMBA, and the chemopreventive agent, genistein.

Summary of Results for Tasks a-c and f-h.

As previously reported and approved in our previous reports, female Sprague Dawley rats were treated prepubertally with genistein and we investigated the proteome in mammary glands of 21 and 50 day old animals (Fig. 1 of Rowell et al. J. Nutri. 135: 2953S-2959S, 2005). Mammary glands were subjected to 2-D gel separation and then the stained gels were scanned *via* a BioRad flat bed densitometer and analyzed with the Progenesis (nonlinear) 2-D gel software system. Differentially regulated protein spots were excised from the gels, destained and prepared for tryptic fragmentation. Tryptic digests of each protein were analyzed using a Voyager MALDI TOF and the spectra produced were matched against a non-redundant protein database using MASCOT to determine the protein's identity. One protein, GTP-cyclohydrolase one (GTP-CH1) was confirmed as being significantly up-regulated at d 21 by immunoblot analysis (Table 1 and Fig. 2 of Rowell et al. J. Nutri. 135: 2953S-2959S, 2005). Investigating down-stream signaling from GTP-CH1, we found that tyrosine hydroxylase was up-regulated and vascular endothelial growth factor receptor two (VEGFR2) was down-regulated in mammary glands of 50 d old rats treated prepubertally with genistein (Fig. 5 and 6 of Rowell et al. J. Nutri. 135: 2953S-2959S, 2005). This and previous work suggest that early prepubertal exposure to genistein enhances cell proliferation by up-regulating GTP-CH1 and the epidermal growth factor (EGF)-signaling pathway, and hence cell differentiation and gland maturation (Fig. 8 of Rowell et al. J. Nutri. 135: 2953S-2959S, 2005). This unique developmental maturation leads to a new biochemical "blue-print" whereby the cells have reduced EGF-signaling and VEGFR2 that render the mature mammary gland less proliferative and less susceptible to cancer.

Tasks d. Using 2-D gels we had determined that guanine deaminase was down regulated in 75 day old rats exposed on day 50 to DMBA. However, this did not confirm by western blot analysis. No change in protein profile was found at day 100.

Task e. No histologic differences were reported in mammary glands of 75 and 100 day old rats treated with DMBA at day 50. This indicates that the DMBA effect is not manifested at the morphology level so soon.

Task i. The 2006 AACR meeting was attended.

Task j. Since our last report a technical paper describing our laboratory and statistical methods was published in Journal of Proteomics and the data on genistein mechanism of action in the mammary was published in Journal of Nutrition. PDF files of both manuscripts are attached.

Specific Aim 2. To collect proteins from interstitial fluid surrounding mammary glands of rats, and to identify and characterize the major proteins that are modulated by DMBA and genistein.

a) Design and build microdialysis probes. At last report, we were using commercially available MF-7026 ultrafiltration probes, with a molecular weight "cut-off" of 30,000 to collect mammary interstitial fluid. However, as there are no commercially available probes that provide membranes with greater than a 30 kDa molecular weight cut-off, we have engineered our own probes.
b-f) Procure and treat animals to use for interstitial fluid collection. This is an ongoing process. However, the progress of these tasks are dependent on the technology that we are trying to optimize.
g) Interstitial fluid dialysates will be subjected to LC-MS/MS for analysis and identification. Using 30 kd cut-off probes, we have become proficient in placement and recovery of only 15-50 ul of fluid each time. However, even more daunting is the very low protein yields from collected samples that are more complicated by dilution problems when we perform downstream separation procedures. Spot intensity from 2-D gels is too weak to proceed to mass spectrometry.
h) Tissue slices will be taken from each mammary gland of 75 and 100 day old rats used in the experiments for histomorphological evaluation. This task is tabled until we succeed with tasks a-g.
i) Attend and present data at AACR meeting. We did attend the 2006 AACR meeting.
j) Data analysis and report to DOD will be written and submitted. Manuscripts will be written and submitted to peer-reviewed journals. While we did publish 2 papers on Aim 1, we do not yet have publishable data from Aim 2.

Previously, we reported the potential of using micro HPLC and commercially available "lab on a chip" for separating proteins. Despite commercial claims, this has proven to be less than successful. Since the 30 kd cut-off probes do not yield sufficient protein to carryout 2-D gel separation and quantitation, and mass spectrometry identification, we have engineered probes that have 400 kda cut-off. We are presently investigating the reproducibility of "hand-building" these probes.

KEY RESEARCH ACCOMPLISHMENTS

- We have developed reproducible 2-D gel and biostatisical methodologies to evaluate the proteome of mammary glands ((Journal of Proteome Research, 4:1619-1627, 2005).
- Prepubertal genistein treatment up-regulates GTP-CH1 in mammary glands of 21-, but not 50-, day old female rats (J. Nutri. 135: 2953S-2959S, 2005).
- Prepubertal genistein treatment up-regulates P-ERK-1 in mammary glands of 21-, but not 50-, day old female rats (J. Nutri. 135: 2953S-2959S, 2005).
- Prepubertal genistein treatment up-regulates tyrosine hydroxylase in mammary glands of 50, but not 21, day old female rats (J. Nutri. 135: 2953S-2959S, 2005).
- Prepubertal genistein treatment down-regulates VEGFR-2 in mammary glands of 50, but not 21, day old female rats (J. Nutri. 135: 2953S-2959S, 2005).
- In the 21 day old animals we found no significant changes in BH4 and iNOS protein expressions (J. Nutri. 135: 2953S-2959S, 2005).

REPORTABLE OUTCOMES

C. Rowell, G. Puckett, K. Roarty, M. Kirk, L. Wilson, M. Carpenter and C. A. Lamartiniere, "Serum profiling and biomarker discover of rat mammary tumors using mass-coded abundance tags

(MCAT)" In Proceedings of the 95th Annual meeting of the American Association for Cancer Research, 45, 1203 Orlando, FL, 2004 .

C. Rowell, and C. Lamartiniere, "Discovery of a Novel Pathway of Chemoprevention by Genistein using Proteomics" Susan G. Komen Mission Conference, New York, NY, 2004

Carpenter, M., Rowell, C, Lamartiniere, C. and McCorkle, H., "2D-gel Proteomics in biomarker discovery." In Proceedings of Pharmaceutical Industry SAS Users Group 2004, San Diego, California

C. Rowell, and C. Lamartiniere, "Proteomic Discovery of Genistein Action in the Rat Mammary Gland. 96th Annual Meeting of American Association for Cancer Research, abstract 398, 2005.

Rowell, C., Carpenter, D.M. and Lamartiniere, C.A. Modeling Biological Variability in 2-D gel Proteomic Carcinogenesis Experiments. Journal of Proteome Research, 4:1619-1627, 2005

Rowell, C., Carpenter, D.M. and Lamartiniere, C.A. Chemoprevention of Breast Cancer, Proteomic Discovery of Genistein Action in the Rat Mammary Gland. J Nutrition Sci, 135:2953S-2959S, 2005.

Lamartiniere, C.A. University of Gottingen, Germany Seminar: Dietary Polyphenols Protect Against Mammary and Prostate Cancers. February, 2006.

Lamartiniere, C.A. Humboldt University, Berlin, Germany Seminar: Dietary Polyphenols Protect Against Mammary and Prostate Cancers. February, 2006.

Lamartiniere, C.A. Technische Universitat, Dresden, Germany Seminar: Dietary Polyphenols Protect Against Mammary and Prostate Cancers. February, 2006.

CONCLUSION

We have developed reproducible methodology for separation, quantitative biochemical analysis, statistical analysis and identification of mammary gland proteins. We have discovered that GTP-CH1 expression is significantly increased shortly following exposure to genistein. However, this was not a permanent effect because by day 50, in the absence of genistein, there was no difference between the treatment groups. By evaluating related metabolic pathways, we have been able to identify down-stream targets and evaluate changes in these proteins in response to the changes of GTP-CH1. In the 21 day old animals we found no significant short-term changes in the tyrosine hydroxylase and iNOS protein expressions. However, at the 50 day time point there was significant up-regulation of tyrosine hydroxylase and VEGF-R2 expression. Since this difference is measurable 30 days after the final genistein treatment, we postulate that some underlying programming effect on protein expression is manifested in the long-term expression profile of this downstream target. This and previous work suggests that early postnatal (prepubertal) exposure to genistein enhances cell proliferation by up-regulating GTP-CH1 and the EGF-signaling pathway, and hence cell differentiation and gland maturation. This unique developmental maturation leads to a new biochemical "blue-print" whereby the cells have reduced EGF-signaling and VEGFR2 that render the mature mammary gland less proliferative and susceptible to chemically induced mammary cancer initiation, angiogenesis and for cancer progression. Therefore, genistein acts through a diverse and coordinated effect of signaling mechanisms and pathways that likely account for the cellular changes responsible for its chemopreventative action. This study demonstrates the usefulness of proteomics for the discovery of novel pathways that may be involved in cancer prevention.

Our experiments in identifying protein biomarkers in interstitial fluid surrounding mammary glands in rats have been much more difficult and not successful to-date. We have recently developed our own probes that should allow us to collect a higher yield of proteins and hopefully allow success.

REFERENCES

1. Lamartiniere, C.A., Cotroneo, M.S., Fritz, W.A., Wang, J. Roycelynn Mentor-Marcel, R. and Elgavish, A. Genistein Chemoprevention: Timing and Mechanisms of Action in Murine Mammary and Prostate. J. Nutrition. 132: 552S-558S, 2002.
2. Shu, X.O. et. al. Soyfood intake during adolescence and subsequent risk of breast cancer amoung Chinese women. Carcinogenesis. 10: 483-488, 2001.
3. Wu, A. H. et. al. Adolescent and adult soy intake and risk of breast cancer in Asian-Americans. Carcinogenesis 23: 1491-1496, 2002.

APPENDICES

Rowell, C., Carpenter, D.M. and Lamartiniere, C.A. Modeling Biological Variability in 2-D gel Proteomic Carcinogenesis Experiments. Journal of Proteome Research, 4:1619-1627, 2005

Rowell, C., Carpenter, D.M. and Lamartiniere, C.A. Chemoprevention of Breast Cancer, Proteomic Discovery of Genistein Action in the Rat Mammary Gland. J Nutrition Sci, 135:2953S-2959S, 2005.

research articles Journal of proteome research

Modeling Biological Variability in 2-D Gel Proteomic Carcinogenesis Experiments

Craig Rowell,[†] Mark Carpenter,[§] and Coral A. Lamartiniere*[,†,‡]

Department of Pharmacology and Toxicology, UAB Comprehensive Cancer, University of Alabama at Birmingham, Birmingham, Alabama 35294, and Department of Mathematics and Statistics, Auburn University, Auburn, Alabama

Received May 3, 2005

We propose a statistical method to model the underlying distribution of protein spot volumes in 2-D gels using a generalized model (GM). We apply this approach to discover mechanisms of chemical carcinogenesis in a rodent model. We generated 247 protein spots that were common to all gels (n = 18). Traditional statistical methods found 6.5% (13 out of 247) significant protein spots, our GM approach yielded a total of 53 (22.5%) differentially expressed protein spots.

Keywords: statistics • 2-D gels • proteomics • carcinogenesis • DMBA • rat

1.0. Introduction

Since the first major studies using two-dimensional gel electrophoresis (2-D gels), the field of proteomics has undergone rapid growth and development.[1] Coupled with mass-spectrometry based protein identification, 2-D gels have been viewed by scientists as a tool for the discovery of proteins and pathways in numerous systems.[2−5] Progress in proteomics research has been directly related to the availability of standard reagents, protocols, and computer programs for data analysis (i.e., Progenesis and PDQuest).[6] These improvements have increased the number of treatment/comparison groups as well as the number of biological replicates within each group that can be examined. In addition, better imaging and processing software allows for attention on proper statistical design and analysis of experiments.

Postrun analysis is the bottleneck of 2-D gel experiments due to high dimensional data likely having high variability.[7] Deficiencies in experimental design and execution greatly impede postrun analysis and decrease the overall sensitivity of the technique. Problems related to analysis first arise in the software processing of the gels, as reported in Nishihara and Champion.[8] These results point to the issue of false positive discovery vs accuracy as a tradeoff affecting the choice of software to use. Another consideration is building composite gels to increase the number of real spots to analyze. Central to composite gel analysis is how to treat absent spots (i.e., averaging intensities vs choosing a "best-of" analysis). Technology such as CyeDyes can potentially overcome this problem, but not without introducing other considerations. Mauer et al. examined 2-D gel data using statistical processes inherent in the analysis software as well as algorithms applied to microarray data.[9] Recently, Chang et al. investigated the issue of spot normalization (a computer generated process) to address the issue of missing values (spots that are represented in the majority but not all of the gels in a data set).[10] A modeling procedure used by Gustafsson et al. adjusts for variances in spot volume data by applying alternative transformations.[11] That each of the above approaches has had a measure of success shows there are numerous approaches for evaluation of 2-D gels.

Much of 2-D gel analysis is based on the search for significant variation between the means (medians) in different groups using the two-sample t-test and analysis of variance (ANOVA). The assumption is that the populations being studied are normally distributed with constant variances, independent of the mean expression levels. If the assumptions are violated, transformations (i.e., log) are taken to make the data more closely conform to the normal distribution. However, this approach has produced limited success because the transformation is usually taken across all analysis variables. Gustafsson et al. noted that even after they transformed their 2-D gel expression data, substantial variance heterogeneity remained.[11] So, rather than manipulating the data until it conforms to preconstructed assumptions, we propose to model the data separately for each protein.

From evolution and development literature we borrow the term "standard norms of reaction" (NoR) to introduce our modeling process of 2-D gel data. Woltereck introduced the concept of NoR to represent the variation of phenotypic response to environmental alterations based on the genotype of the organism.[12,13] The environmental condition in our current study is the process of carcinogenesis. In this study the same genetic strain of animals has been exposed to the same environmental insult (dimethylbenz[*a*]anthracene, DMBA). We know that this experiment will result in the production of mammary tumors in all treated animals. We also know that the timeline of palpable tumor development is variable among

* To whom correspondence should be addressed. E-mail: Coral@uab.ed.
[†] Department of Pharmacology and Toxicology, University of Alabama at Birmingham.
[‡] UAB Comprehensive Cancer, University of Alabama at Birmingham.
[§] Auburn University.

10.1021/pr0501261 CCC: $30.25

Published on Web 08/13/2005

the individual animals; therefore, there is an underlying plasticity in the phenotypic response.[14,15] To avoid confounding effects of tumor heterogeneity, we will look at the period of early lesion formation.[14] In general, we presume that changes observed at this time point will reflect early biochemical events related to promotion. It is our goal to model a tissue protein signature(s) associated with early cancer formation.

In this paper, we describe the importance of statistical design and analysis when conducting investigations using 2-D gels for differential protein expression profiling. A series of experiments and analyses related to our research into the biochemical mechanisms of carcinogenesis by DMBA in a rodent mammary model provide the data. We propose a statistical method whereby the underlying distribution of spot volume is modeled directly as a generalized distribution. This generalized model (GM) encapsulates the various methods of transformations and analyses found in modern proteomic literature. The GM method will therefore yield better rates of discovery than more traditional proteomic statistical analyses and better reflect biological changes in protein expression.

2.0. Materials and Methods

Sprague–Dawley CD rats were purchased from Charles River Breeding Laboratories (Raleigh, NC). Dimethylbenz[*a*]anthracene (DMBA) and sesame oil were purchased from Sigma Chemical Company (St. Louis, MO). Isoelectric focusing (IEF) strips, IEF buffer, Multiphor II, tissue grinding kits, and albumin removal kits were purchased from Amersham Biosciences (now a member of GE Healthcare, Piscataway, NJ). All other chemicals were purchased from Fisher Scientific (Hampton, NH). SyproRuby and the VersaDoc densitometer were purchased from Bio-Rad (Hercules, CA). SAS v.10 was purchased from the SAS Institute (Cary, NC).

2.1. Pilot Projects, Replication and Power Analysis. One important aspect of experimental design is choosing sample size.[16] In this study, we promote the use of power analysis in determining sample size. Other issues of statistical design are the elimination of extraneous sources of variability and choosing the number and levels of comparison groups. Our first consideration is the choice between technical and biological replications.

Technical versus Biological Replication. Using technical (analytical) replicates over biological replicates has been widely discouraged.[17–19] However, Asrivatham et al. stated that the use and investment in analytical replicates for pilot projects is extremely valuable for data quality control and validation of the 2-D gel handling process.[20] Importantly, the general consensus is that as functions of cost and resources, biological replicates provide considerably more scientific information than analytical replicates.

Power and Sample Size. Pilot studies were conducted to determine optimal sample size based on power analysis. In our study, the power estimate and sample size determinations involved using unique uterine samples (biological replicates) from 8 control- and 8 genistein- (a phytoestrogen found in soy) treated rats. The variance for each of the commonly detected proteins was estimated using the pilot expression data. The variance estimate was used to evaluate sample size effects for discovering specific protein volume fold-changes. Rather than basing power analysis on crude family wise adjustments, such as Bonferroni, we designed an experiment with sufficient power to examine at least one single protein comparison (in our study, the power analysis was based on adjustment of 100 proteins) and after the data was collected we computed the estimated false discovery rate to assess the potential number of discoveries.

False Discovery Rates (FDR). Benjamini and Hochberg first coined the phrase "false-discovery rate" (FDR) now commonly applied in significance testing designed for high dimensional biology.[21–25] For a particular experiment, the FDR is the expected or estimated proportion of false discoveries out of the total number of significantly different genes/proteins. This means that a large FDR of 50% would lead the researcher to a different decision with respect to allocation of resources than if the FDR were 5%. Therefore, we computed the estimated FDR to assess the potential number of discoveries after the data was collected.[23–24]

2.2. Study Design. Animal care and treatment were performed according to established guidelines approved by the UAB Animal Care Committee. Eighteen 50-day-old female Sprague–Dawley rats were divided into two groups and either gavaged with 40 μg DMBA/g B. W. ($n = 8$) or gavaged with an equal volume of vehicle, sesame oil only ($n = 10$). At 75 days of age (25 days post DMBA treatment), animals were anesthetized with Ketamine/xylazene and the fourth abdominal mammary glands were dissected. We selected 75 days post DMBA with the intention of investigating mammary glands with early preneoplastic lesions and biochemical alterations, and yet relatively tumor mass free. Each gland was cut in half longitudinally to allow both proteomic as well as pathological evaluations. Frozen mammary tissues were homogenized in lysis buffer formulated for 2-D gels using tissue grinding kits.[26] After measuring protein concentration via Bradford's assay (Bio-Rad), equal concentrations of sample were subjected to albumin removal. Protein concentration was remeasured and 150 μg protein aliquots were diluted in rehydration buffer. The samples were applied to separate immobilized pH gradient (IPG) strips (24 cm, pH 4–7) and allowed to rehydrate overnight at room temperature. The IPG strips were placed on a flatbed electrophoresis unit (Multiphor II) and a current gradient applied (500 V for 1 h, 3500 V for 1.5 h, followed by 3500 V for 22.5 h). After isoelectric focusing, IPG strips were equilibrated first in 100 mM dithiothreitol for 45 min followed by equilibration in 120 mM iodacetimide for 45 min. IPG strips were loaded onto pre-cast 1.5 mm, 12.5% SDS gels and run on a Dodecacell vertical electrophoresis unit according to manufacturer's suggestions. Both IEF and SDS gels were run as block groups consisting of equal treatments per run. Once gels were run to completion, they were stained using SyproRuby and scanned via a VersaDoc 4000 Densitometer. Spot matching and gel warping were done using Progenesis Discovery 2004. Processed data was imported into SAS version 10 and analyzed using statistical methods and algorithms based on various SAS procedures.

2.3. Data Processing. For our experiments, we elected to use the "total spot volume" normalization procedure found in the Progenesis software. After spot matching and gel warping were completed, the data file was exported to SAS for processing. For all the following procedures, we evaluated only spots that were common to all gels in the data set. The first step in our cleanup procedure was to perform a t-test. Each spot identified as significant ($p < 0.05$) was located and the spot's presence was visually confirmed in all the gels. As needed, manual rematching of spots was conducted and the statistical program was re-run to generate a new list of p-values for the matched spots. This iterative process was run numerous times to ensure

that matches reflected high quality spots (i.e., consistent shape, nonsaturation and proper splitting).

2.4. Statistical Methods. First, we applied traditional statistical approaches to differential expression analysis and their adaptations to assumption violations. Second, since in proteomic studies it is not uncommon to come across data that are nonnormally distributed and/or differently dispersed, we discuss two different ways of dealing with these situations. In section 2.4.2., we describe an approach that we refer to as an indirect method where traditional statistical analysis is conducted on the transformed data. In Section 2.4.3., we describe our direct approach, where general classes of distributions, (generalized gamma, exponential, or Weibull) are directly fitted to the data using a generalized linear model.

2.4.1. Differential Protein Expression. For a given 2-D gel experiment, proteomic differential expression analysis describes the process of conducting multiple hypothesis tests, one for each protein, across all commonly expressed proteins. In the traditional two-sample t-test, any protein resulting in a p-value that is less than a pre-specified α (i.e., 0.05) is considered significant and that protein is deemed differentially expressed. This approach must be implemented with caution, because the error rate is fixed only for one specific test and if more than one hypothesis/protein is tested then the error rate accumulates across all tests.

Since many experiments involve the comparison of two treatment groups and since the approach can be easily generalized to more than two-groups, we focus our attention on the two-population comparison. If there are n_1 and n_2 gels processed in groups 1 and 2, respectively, and the populations are assumed to have approximately equal variances, then the two-sample t-test involves the computation of the following test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2(1/n_1 + 1/n_2)}},$$
$$s_p^2 = [((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/(n_1 + n_2 - 2)] \quad (1)$$

where $\bar{x}$, $\bar{x}_2$, s^2, and s_2^2 are the sample means and variances from each sample, respectively. In the two-sample t-test, if the normality assumption is reasonable but the common variance assumption is violated then eq 1 may not be valid. However, approximate t-tests are available to test for differences between the two means. Cochran and Cox proposed an approximate t-test, but the degrees of freedom were undefined when the sample sizes were unequal and the test was quite conservative.[27,28] Satterthwaite's approximation for the degrees of freedom can be used for the approximate t-test in these cases, but the test given below, still remains conservative.[29,30]

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_1^2/n_2}}, df = (w_1 + w_2)^2/(w_1^2/(n_1 - 1) + w_2^2/(n_2 - 1)), w_1 = s_1^2/n_1, w_2 = s_2^2/n_2)^{29-30} \quad (2)$$

Regardless of the test used (eqs 1 or 2) there is a question of efficiency since one usually tests the equal variance assumption before deciding between the t-test (equal variances) or the approximate t-test (unequal variances). Nonparametric approaches do not provide much relief since they typically assume symmetrically distributed populations with common variance or similar shapes across groups.

Since a 2-D gel experiment involves several hundred hypothesis tests on unknown proteins, it is impossible to know whether the equal variance assumption is true for all proteins. On the basis of the sample expression between two groups, one can test whether the underlying populations have the same spread, dispersion or variance ($\sigma_1^2 = \sigma_2^2$versus $\sigma_1^2 \neq \sigma_2^2$) using the Folded Form F-test (eq 3)

$$F = \max(s_1^2, s_2^2)/\min(s_1^2, s_2^2) \quad (3)$$

which, under the normal distribution assumption, has an F_{n_1-1,n_2-1} distribution. Although most researchers conduct such a test to determine whether the two-sample t-test based on equal variance is valid, we propose that proteins that have significantly different variances between groups may well be of biological significance. That is, if a protein has significantly different variances between groups, then it is included in the list of significant proteins whether there are significant mean differences in expression.

2.4.2. Transformation Approach. In genomic and proteomic studies statistical analyses is often conducted on the log-transformed data across all genes and proteins. In many cases, this approach results in more symmetrically distributed data and/or dampens the effect of nonconstant variance at high levels and outliers. However, Rocke, and Durbin provided evidence that for low expressing genes or proteins, this transformation can make matters worse.[31,32] Accordingly, much literature has been dedicated to more generalized transformations such as Box-Cox[33] and Generalized-log transformations that serve as an alternative to blind application of a single transformation.[32,34] Specifically, if y represents the expression value for a particular protein or gene, then the simple Box-Cox transformation is of the form $z = (y^\lambda - 1)/\lambda$ if $\lambda \neq 0$ and $z = \log(y)$ if $\lambda = 0$.[33] This class includes most of the common transformations, including the log-transform and various power and inverse power transforms. The underlying goal in using a generalized transformation is that the resulting data will be more in line with the model assumptions and therefore produce more robust analyses of the data.[31,32,34-38] The generalized class of transformations are appealing because they are very flexible. They include a form of the simple log-transform as a special case, and the appropriate transformation can be estimated using maximum likelihood approaches.[35] The TRANSREG procedure in SAS offers the maximum likelihood approach in fitting the optimal Box-Cox transformations to data taken from Draper and Smith.[39] The model fitting feature allows one to optimize and/or customize the transformation for each individual protein or gene rather than doing a single log-transform across all proteins or genes.

2.4.3. Generalized Model (GM). The generalized model more directly addresses the problems discussed above by providing a unified theoretical and conceptual framework for analyzing protein differential expression across each protein spot. Generalized models assume the response variable (expression) is not necessarily normally distributed and the underlying distributions may not have constant variances between groups or across levels of the predictor variables.[40] In many cases other than the normal distribution, the populations may have a mathematical dependency (link function) between the variance and the mean of the populations. The GENMOD procedure in SAS provides Newton–Raphson algorithm (ridge-stabilized) to maximize the log-likelihood function in estimation and testing of parameters in the model for a broad collection of models, including the normal, inverse-Gaussian, gamma, negative binomial, and Poisson distributions.

Therefore, we propose a method for 2-D gel analysis whereby the underlying distribution is modeled directly as a generalized-gamma distribution, which has the Weibull, exponential, gamma, and log-normal as special cases. Each of these special distributions has a relationship with a log-location-scale family of distributions. For example, taking the log-transformation of Weibull, gamma, and log-normal data leads to the extreme value distribution, the log-gamma distribution and the normal distribution, respectively. Each of these distributions is a special case of the generalized log-gamma distribution.[40] Therefore, under the right conditions, fitting the generalized gamma or the generalized log-gamma distribution to data leads to distributions approximating the true underlying distributions individually and perhaps more accurate statistical contrasts between treatment groups. Inference can then be made about the location, shape and scale of the distribution without having prior knowledge of the specific positive support distributions across all proteins within the given populations or treatment groups. Therefore, when the goal is discovery of proteins, we propose a method where the generalized gamma distribution is fit to each specific commonly expressed protein within populations and tested for significant differences across the populations. The new list of proteins is then compared and contrasted to those found as worthy of follow-up analysis through other more traditional methods, including tests on mean and variance differences.

Our GM method is expressed as follows: Y denotes the expression for a particular spot on a 2-D gel, and Y has a generalized gamma distribution if its distribution function is of the form

$$f(y) = \frac{|\delta|}{y \cdot \Gamma(\delta^{-2})}(\delta^{-2}y^{\delta})^{\delta^{-2}} \exp(-\delta^{-2}y^{\delta}), y > 0 \qquad (4)$$

where $\Gamma(\cdot)$ is defined as the gamma function. Taking the log-transform of a generalized-gamma random variable, $z = \log(y)$ results in the location-scale family called the generalized-log-gamma distribution, given in its standard form as follows:

$$f(z) = \frac{|\delta|}{\Gamma(\delta^{-2})}(\delta^{-2}\exp(\delta \cdot z))^{\delta^{-2}} \exp(-\delta^{-2}\exp(-\delta \cdot z)),$$
$$z, \delta \in (-\infty, \infty) \quad (5)$$

The parameter δ is referred to as the shape parameter. If $\delta = 1$, then the log-generalized gamma becomes the extreme value distribution and the corresponding generalized gamma becomes the Weibull distribution. If $\delta = 0$, then the log-generalized gamma becomes the normal distribution and the corresponding generalized gamma becomes the log-normal distribution. Regression analysis based on these models can be done by using the LIFEREG, NLP, or NLIN procedures in SAS. The typical approach is to log-transform the data first, and then fit the generalized log-gamma distribution separately to each of the protein expression variables, which is equivalent to fitting the corresponding generalized gamma to the raw data. The density in eq 5 is expressed in standard form (just as a normal distribution with mean zero and unit variance is the standard form of the normal family). As Lawless pointed out, the generalized log-gamma (GLG) distribution is a location-scale family, just like the normal family, and these parameters are introduced into the density by letting $z = (u - \mu)/\sigma$, and u becomes a GLG (μ,σ,δ), where μ, σ, and δ represent the location, scale and shape parameters, respectively.[41]

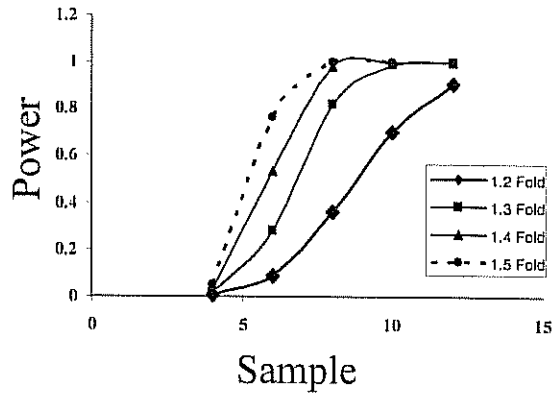

Figure 1. Power analysis versus sample size. This graph illustrates how power and sample size are related with respect to detection of fold change in protein expression.

The location can be expressed in terms of linear regression model on the log-transformed data. Initial estimates of regression parameters are obtained by doing ordinary least-squares regression on the log-transformed data, which are then used to get more precise maximum-likelihood estimators (MLE) using some numerical method such as ridge-stabilized Newton−Raphson algorithm. Differences in location between populations can then be tested directly using a χ-square test (Wald test). The GENMOD or NLP procedures could be used to generalize this approach for simultaneously testing for differences in location (mean/median), scale (variances/standard deviation) and shape, but for illustrative purposes in this paper we focus on tests for differences in location in models in which the mean and variance are mathematically related.

Results

3.1. Power Analysis/Sample Size Determination. In our first pilot study, five replicate 2-D gels from a single uterine sample were used to examine reproducibility. The results showed that the total number of protein spots per gel were reasonably similar. However, when we looked only at common spots among the gels, we found that as the multiplicity of gels increased there was a significant decrease in number of common spots (unpublished data). This is consistent with earlier findings reported by Voss and Haberl.[7]

A second pilot study using uteri from 8 control- and 8 genistein-treated rats also showed that as the multiplicity of gels increased there was a significant decrease in number of common spots. Since this decrease in matched spots was at similar rates between groups, it indicated a lack of sample handling bias. The pooled control estimate of standard deviation in normalized peak intensities was used to determine that a sample size of 8 animals per treatment group would be sufficient to detect a 1.5-fold-change between the two groups. This change was detected with over 99% power, based on a two-sample t-test with an experiment-wise level of significance of $p < 0.05$, with adjustments for multiple testing. Figure 1, displays the power curves for the detection of four different fold changes (1.2, 1.3, 1.4, and 1.5). The power is defined as the probability of detecting the specified fold-change and is displayed over sample sizes ranging between 4 and 12. While a sample size of 8 gives 99% power to detect a 1.5-fold-change, this power drops to 82% to detect a 1.3-fold-change. A sample size of 6 only gives 28% power to detect a 1.3-fold-change.

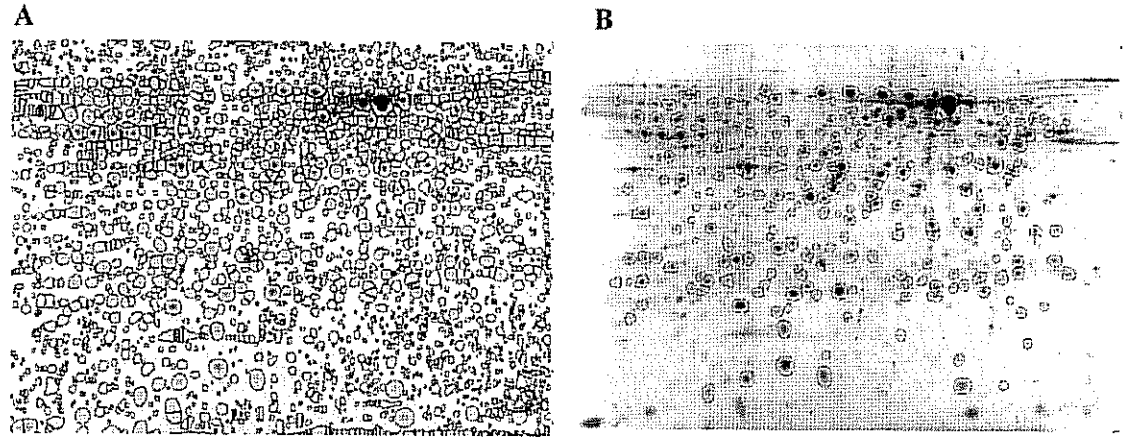

Figure 2. 2-D gel profile (A) A display of unsupervised spot detection results. (B) A display of those spots common to all gels in the experiment.

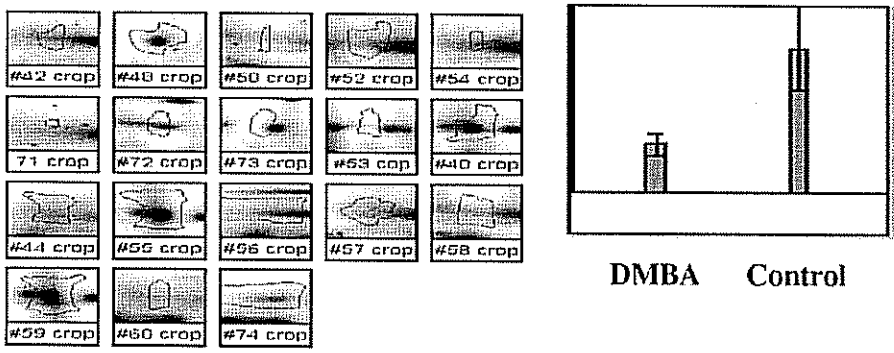

Figure 3. Supervised spot evaluation. Initial evaluation of common spots is based on the *p*-value from a two-sample t-test. The graph shows the mean value (±SEM) for the normalized value. Spots are ranked according to *p*-values and all spots with a $p < 0.05$ are subjected to visual inspection to verify consistency of spot parameters. This figure shows that while the t-test on the normalized value was significant ($p = 0.0431$), there is inconsistency in spot detection.

3.2. Data Quality and Processing. Results of unsupervised matching and spot detection (Figure 2A) demonstrate the need for a directed process of image cleanup before evaluation. After initial matching we focused only on those spots found in all gels (Figure 2B). Common spots with significant *p*-values ($p < 0.05$, t-test) were subjected to visual verification to ensure both accuracy of matching and consistency of spot boundary (Figure 3). This early analysis is critical to prevent improper data interpretation. Once several new landmarks have been established the matching program and inspection process is rerun. This iterative process greatly increases the efficiency of subsequent evaluations by providing well matched data points for the more robust statistical procedures.

3.3. Statistical Analysis of Two Experimental Groups. Our primary data set was generated using 18 gels representing unique mammary gland samples in each of two treatment groups (10 control and 8 DMBA treated rats). Analysis of all 18 gels yielded 247 spots that were present in every gel. These 247 common spots were subjected to statistical differential expression analysis. Evaluation of the data using only the t-test on the untransformed data found 13 spots to be significantly different between the 2 groups ($p < 0.05$) (Figure 4A). Testing of the log-transformed data yielded a total of 15 spots to be significantly different ($p < 0.05$) (Figure 4B). GM calculations added an additional 11 spots for a total of 26 spots that were significantly different ($p < 0.05$) (Figure 4C). Using a 0.05 level of significance, the estimated FDR was 0.20. Therefore, we expected 5 of the 26 spots found using the GM to be false positives.

3.3.1. Generalized Models. An advantage of the GM procedure is that it allows for the mean and variance to be linked and vary simultaneously between groups. Individual data plots for three spots where the *p*-values differed for the t-test, log-normalized and GM are presented in Figure 5. For each protein spot the individual data points of mammary glands of control- and DMBA-treated animals are graphed to show the variation for the log-normalized data. For those instances where the norm and log-normalized data are not significantly different we assume that the mean values are similar. Graphs in Figure 5A,B demonstrate that while the means are similar, the underlying variation of expression is different. Therefore, using the GM we model this variation and determine the spots to be significantly different ($p < 0.05$).

3.3.2. Tests on Equal Variances (Folded Form F-test). Figure 6 illustrates that using just the Folded Form F-test (testing only on variance) we find 33 unique proteins not captured by any of the other tests. Finally, we see that there is an overlap of only 3 spots identified as being significant using all testing procedures. The field graph in Figure 7 illustrates all 247 protein spots that were evaluated. This graph reveals the overlap of significantly evaluated spots. Each significant spot's location is based on either differences only in the variance as a function

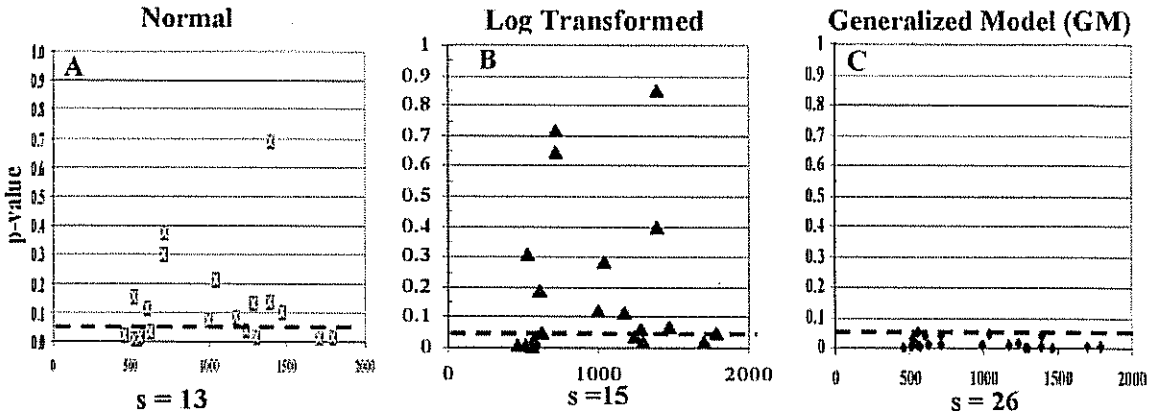

Figure 4. Comparison of traditional and GM testing procedures. The number of significant spots for each testing procedure is represented below the dashed line ($p < 0.05$). (A) The t-test applied to the normalized data found 13 spots to be significantly different. (B) Means testing on the logtransformed data found 15 spots differentially expressed and (C) results using the GM captured 26 spots as differentially regulated.

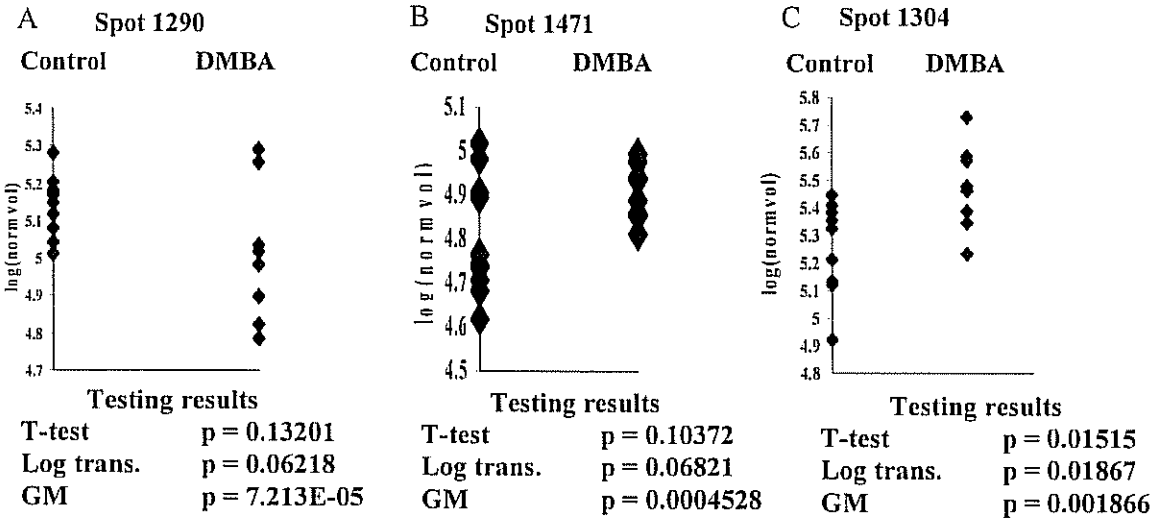

Figure 5. Consideration of variance. Each graph displays the log-transformed data for an individualsample in either group (Control or DMBA) (A) Results of the two-sample t-test on the normal or logtransformed data for spot 1290 are not significant ($p > 0.05$). However, results of the GM show a highly significant ($p = 7.213 \times 10^{-05}$) difference in variance between the control and DMBA groups. (B) For Spot 1471 results of the two-sample t-test on the normal or log-transformed data are not significant ($p > 0.05$). However, the GM found a highly significant ($p = 0.000\ 452\ 8$) difference in variance between the control and DMBA groups. (C) For spot 1304, the results of the two sample t-test were significant and with log transformation ($p < 0.05$), as well the GM was significant ($p = 0.001\ 866$).

of the same mean, or variance in the absences of similar means. Finally, spots uniquely found significant using the GM procedure are distinguished in the broad field.

4. Discussion

Proteomics and genomics fall under the general heading of systems biology. Systems biology focuses on the interaction of all molecular components including: DNA, RNA, proteins, protein interactions, biomodules, cells, tissues, etc., with each of these components having their own individual elements (e.g., specific gene methylation or protein post-translational modifications). A systems level view is necessary to understand the complex dynamics that underlie the physiology in both the normal and diseased states. Systems biology is characterized by a synergistic integration of theory, computation, and experiment.[42]

Advances in recent technology make possible the large-scale application of proteomics for biomarker discovery in cancer models and the exploration of mechanisms of action of drugs. These advances result in the ability to readily run reproducible 2-D gels for protein separation and obtain protein identification using mass spectrometry techniques, such as MALDI-TOF. Software programs, such as Progenesis, have been developed that aid the researcher in evaluating changes in protein expression profiles among groups and between samples. However, these programs lack substantial statistical analysis tools to help researchers determine the most important and persistent changes throughout the experiment. Without adequate means of analysis the researcher is left to generate a long list of proteins for identification, and then is required to use a hit-or miss strategy for further analysis.

The 2-D gel cleanup/spot review and evaluation cycle has long been considered the bottleneck of 2-D gel experiments. This has resulted from over reliance on the unsupervised matching and spot evaluation by the software followed by an unscripted procedure for cleanup by the end-user. Therefore,

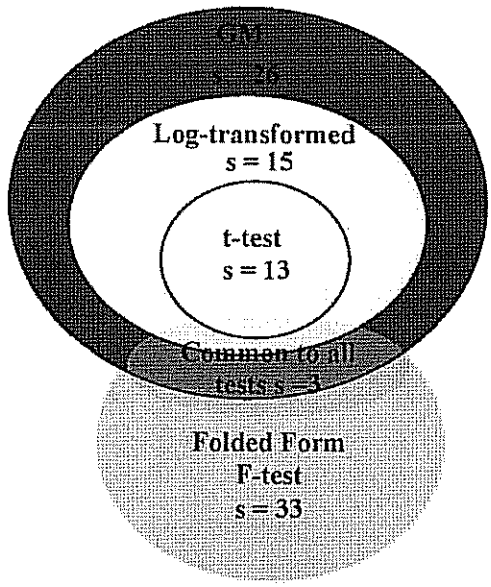

Figure 6. Nested collection of significant spots. The GM procedure found the same spots as the traditional t-test as well as those found from testing on the log-transformed data. The number of unique spots using the Folded Form F-test ($s = 33$) are demonstrated by the green circle. Three spots were found to be significant regardless of testing method.

we have developed a method that greatly increases the speed of this process by providing guidance and direction. Through multiple trials we have determined that statistical analyses are best conducted only on the common spots.

Our current research focuses on finding biochemical events that indicate the earliest stages of breast cancer development. Using an animal model of carcinogenesis, we developed our evaluation of markers along a known timeline of tumor development. The DMBA model we chose has been demonstrated to result in 100% mammary tumor incidence. In general, we have seen that DMBA treatment to 50 day old rats results in palpable tumor development when the animals are 100–120 days old; therefore our choice to evaluate mammary glands at 75 days of age (25 days after DMBA administration) represents a very early state of carcinogenesis. Pathological examination of these animals showed no lesion formation in the DMBA treated animals at day 75. Given that cancer is a disease process with a long developmental period we acknowledge that the earliest stages of carcinogenesis are likely marked by subtle alterations in protein expression. These low expression differences are one reason that we have emphasized power analysis to provide information about our lower limits of detection in 2-D gel experiments.

Power analysis is a standard method to determine a level of sensitivity for value change (such as spot volume fold change) as a function of the sample size. In any biomedical experiment, the number of experimental units (sample size) should be selected to maximize the probability (power) of detecting a predetermined significant difference between two or more treatments (i.e., protein fold change). By addressing the issue of sensitivity from the beginning, this knowledge can be applied to help determine if the changes in expression of a particular protein make logical sense for the given experimental design/biology. While replication studies for power determination can be costly, establishment of statistically relevant data will lead to reduced end-cost. For our biological model, the result of power and sample size determination established our ability to confidently identify those spots that differed in mean expression by 1.5-fold or greater with a reasonable number of biological replicates. However, results of traditional expression evaluation, t-test and log transformed data, only identified a finite number of significant spots ($s = 13$ and $s = 16$, respectively). In fact, finding 13 to 16 spots represents only 5–6% of the total evaluated protein spots ($s = 247$). This low value, while technically accurate, represents a level of finding

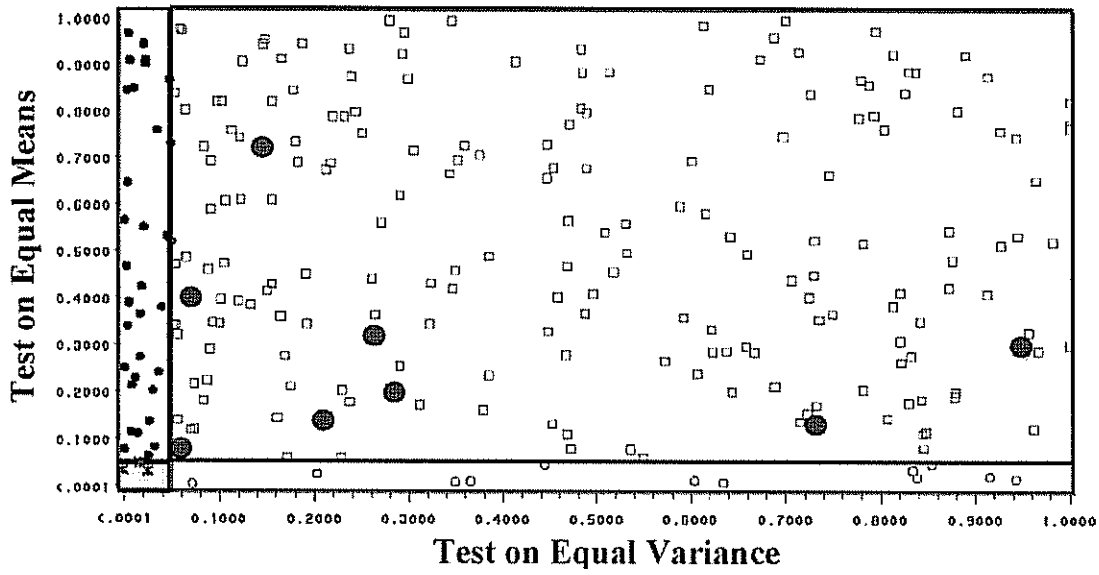

Figure 7. Field graph of commonly expressed spots. The vertical axis represents the p-values from a two-sample t-test conducted on the log-transformed data (shaded area represents $p < 0.05$). The horizontal axis represents the p-values resulting from a test on equal variances among the groups (shaded area represents $p < 0.05$). Red circles depict those p-values based on the generalized linear model (GM) that were significant.

not unlikely to be based on chance. Therefore, we needed to design a more robust approach to evaluate our data.

As mentioned in the Introduction, the application of the GM provides additional information on the distribution of the individual data points for a particular spot. By applying the concept of NoR to our evaluation we saw alteration in the variance levels, either tighter regulation or dysregulation for some of the proteins examined, while the mean appears similar (see Figure 5). Spread of variation shows the natural characteristic of the model to allow for wide fluctuations in the normal circumstance, or the inverse that certain proteins require strict control to maintain adequate cellular function. By evaluating the changes in variance of expression we gain insight into a level of control that may be involved in the promotion of carcinogenesis. Since the result of 2-D gels is to look at a broad spectrum of proteins, we may be able to establish patterns of variance alteration and determine if proteins that undergo positive or negative shifts in expression are functionally related to one another or the disease process.

The changes of any particular protein over the course of tumor development will itself alter as, in the case of mammary cancer, the underlying cell population changes.[14] Traditionally, tumorogenesis is measured as a mean time to tumor development, hence we have to use multiple animals/group to get mean to first, second, etc. tumor/rat since the individual animal's response is different. Furthermore, it is the fundamental effect of treatments such as cancer promoters or chemopreventive agents to alter the time of tumor development. However, all of these measures ignore the individual response or the general group response unless the mean levels are significantly different. Ultimately, we appreciate that underlying alterations involved in the long term process of carcinogenesis will likely be found in subtle, yet persistent, changes in cellular signaling.

It is well recognized that the value for an individual spot on a 2-D gel does not necessarily represent an absolute measurement for the concentration of a protein. For this reason, we acknowledge that there is some inherent weakness in performing exhaustive evaluations of spots from a statistical standpoint. It is our assumption that investigators are willing to make certain tradeoffs in data quality vs time and future evaluation. That is to say, any mass-spectrometry based protein identification is going to require more stringent confirmation procedures, such as immuno techniques. In turn, these techniques will allow for a more quantitative assessment of changes in protein concentration. It is our intent to provide more information about the general qualities of the information that the 2-D gel is providing and to help guide the researcher in the decision making process with respect to which spots should be evaluated first. Therefore, displays such as Figure 7 provide all of the information with respect to what model resulted in a spot being found significant. This system alleviates the production of laundry-lists of proteins and allow for directed and focused studies of particular proteins/pathways that are involved in the condition under study. Therefore, our future experiments will be designed to more accurately capture data related to the temporal changes we have observed to better establish the role of identified proteins.

In summary, we have described a reproducible and statistical approach to the use of 2-D gels for identification of biomarkers that may be related to the carcinogenesis of DMBA in the rat mammary gland. These methods lend well to the discovery of novel new proteins and identification of key signaling pathways involved in cancer causation. Our statistical approach involves empirical determination of the number of gels required to ensure statistical power for appropriate evaluation. In general, the approach we used results in quickly identifying those proteins that meet a realistic and significant change, but is also broad enough to allow the unique modeling approach of the GM. The approach that we have outlined is what we consider to be discovery proteomics. Only when we have mass spectrometry data for identification do we consider this as our preliminary data, not as conformational or primary data. Experiments can then be designed to evaluate the validity of identifications including the previous mention of more specific techniques of quantification.

Acknowledgment. C.R. was supported by stipends from DOD BC 17-00-1-0119 and the Susan G. Komen Foundation DISS0201242. This research was funded by Susan G. Komen Foundation DISS0201242, NIH 1 R01 CA61742-07, NIH 1 UO1 ES012771-01, and DOD BC 021371.

References

(1) O'Farrell, P. Z.; Goodman, H. M. Resolution of simian virus 40 proteins in whole cell extracts by two- dimensional electrophoresis: heterogeneity of the major caspid protein. *Cell* **1976**, *9*, 289–298.

(2) Illsley, N. P.; Lamartiniere, C. A.; Lucier, G. W. Analysis of the sex specific changes in rat hepatic cytosol protein patterns using two-dimensional gel electrophoresis. *J. Appl. Biochem.* **1979**, *1*, 385–395.

(3) Arnott, D.; O'Connell, K. L.; King, K. L.; Stults, J. T. An integrated approach to proteome analysis: identification of proteins associated with cardiac hypertrophy. *Anal. Biochem.* **1998**, *258*, 1–18.

(4) Lewis, T. S.; Hunt, J. B.; Aveline, L. D.; Jonscher, K. R.; Louie, D. F.; Yeh, J. M.; Nahreini, T. S.; Resing, K. A.; Ahn, A. G. Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol. Cell.* **2000**, *6*, 1343–1354.

(5) Hondermarck, H.; Vercoutter-Edouart, A. S.; Révillion, F.; Lemoine, J.; El-Yazidi-Belkoura, I.; Nurcombe, V.; Peyrat, J.-P. Proteomics of breast cancer for marker discovery and signal pathway profiling. *Proteomics* **2001**, *1*, 1216–1232.

(6) Taylor, C. F.; Paton, N. W.; Garwood, K. L.; Kirby, P. D.; Stead, D. A.; Yin, Z.; Deutsch, E. W.; Selway, L.; Walker, J.; Riba-Garcia, I.; Mohammed, S.; Deery, M. J.; Howard, J. A.; Dunkley, T.; Aebersold, R.; Kell, D. B.; Lilley, K. S.; Roepstorff, P.; Yates, J. R. 3rd; Brass, A.; Brown, A. J.; Cash, P.; Gaskell, S. J.; Hubbard, S. J.; Oliver, S. G. A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat. Biotechnol.* **2003**, *21*, 247–254.

(7) Voss, T.; Haberl, P. Observations on the reproducibility and matching efficiency of two-dimensional electrophoresis gels: consequences for comprehensive data analysis. *Electrophoresis* **2000**, *21*, 3345–3350.

(8) Nishihara, J. C.; Champion, K. M. Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* **2002**, *23*, 2203–2215.

(9) Maurer M. H.; Feldmann, R. E.; Bromme, J. O.; Kalenka, A. Comparison of statistical approaches for the analysis of proteome expression data of differentiating neural stem cells. *J Proteome Res.* **2005**, *4*, 96–100.

(10) Chang, J.; Van Remmen, H.; Ward, W. F.; Regnier F. E.; Richardson, A.; Cornell, J. Processing of data generated by 2-dimensional gel electrophoresis for statistical analysis: missing data, normalization, and statistics. *J. Proteome Res.* **2004**, *3*, 1210–1218.

(11) Gustafsson, J. S.; Ceasar, R.; Glasbey, C. A.; Blomberg, A.; Rudemo, M. Statistical exploration of variation in quantitative two-dimensional gel electrophoresis data. *Proteomics* **2004**, *4*, 3791–3799.

(12) Woltereck, R. Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer Artunterschiede bei Daphnien. *Verh. Deutsch. Zool. Gesellsch.* **1909**, *19*, 110–173.

(13) Sarkar, S.; Fuller, T. Generalized norms of reaction for ecological developmental biology. *Evol. Dev.* **2003**, *5*, 106–15.

(14) Russo, J.; Wilgus, G.; Russo, I. H. Susceptibility of the mammary gland to carcinogenesis: I Differentiation of the mammary gland as determinant of tumor incidence and type of lesion. *Am. J. Pathol.* **1979**, *96*, 721−736.
(15) Lamartiniere, C. A.; Cotroneo, M. S.; Fritz, W. A.; Wang, J.; Mentor-Marcel, R.; Elgavish, A. Genistein chemoprevention: timing and mechanisms of action in murine mammary and prostate. *J. Nutr.* **2002**, *132*, 552S−558S.
(16) Muller, P.; Parmigiani, G.; Christian, R.; Rousseau, J. Optimal sample size for multiple testing: The case of gene expression microarrays. *J American Stat Soc.* **2004**, *99*, 990−1001.
(17) Simon, R. M.; Dobbin, K. Experimental design of DNA microarray experiments. *Biotechniques Suppl.* **2003**, 16−21.
(18) Yang, Y. H.; Speed, T. Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* **2002**, *3*, 579−588.
(19) Molloy, M. P.; Brzezinski, E. E.; Hang, J.; McDowell, M. T.; VanBogelen, R. A. Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics* **2003**, *3*, 1912−1919.
(20) Asirvatham, V. S.; Watson, B. S.; Sumner, L. W. Analytical and biological variances associated with proteomic studies of Medicago truncatula by two-dimensional polyacrylamide gel electrophoresis. *Proteomics* **2002**, *2*, 960−968.
(21) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., Series B* **1995**, *85*, 289−300.
(22) Allison, D. B.; Gadbury, G. L.; Heo, M.; Fernández, J. R.; Lee, C.-K.; Prolla, T. A.; Weindruch, R. A mixture model approach for the analysis of microarray gene expression data. *Comput. Stat., Data Anal.* **2002**, *39*, 1−20.
(23) Storey, J. D. A direct approach to false discovery rates. *J. Royal Stat. Soc. Ser. B* **2002**, *64*, 479−498.
(24) Storey, J. D. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Stat.* **2003**, *31*, 2013−2035.
(25) Liao, J. G.; Lin, Y.; Selvanayagam, Z. E.; Shih, W. J. A mixture model for estimating the local false discovery rate. *Bioinformatics* **2004**, *20*, 2694−2701.
(26) Fountaoulakis, M, Schull, E, Hardmeier, R, Bernt, P and Lubec, G. Rat brain proteins: Two-dimensional protein database and variations in the expression level. *Electrophoresis* **1999**, *20*, 3572−3579.
(27) Cochran, W. G.; Cox, G. M. *Experimental Designs*; John Wiley & Sons: Inc.: New York, 1950.
(28) Lee, A. F. S.; Gurland, J. Size and power of tests for equality of means of two normal populations with unequal variances. *J. Am. Stat. Assoc.* **1975**, *70*, 933−941.
(29) Satterthwaite, F. W. An approximate distribution of estimates of variance components. *Biometrics Bull.* **1946**, *2*, 110−114.
(30) Freund, R.J.; Littell, R. C.; Spector, P. C. *SAS System for Linear Models*; SAS Institute Inc.: Cary, NC, 1986.
(31) Rocke, D. M.; Durbin, B. A model for measurement error for gene expression arrays. *J. Comput. Biol.* **2001**, *8*, 557−569.
(32) Rocke, D. M.; Durbin, B. Approximate variance-stabilizing transformations for gene-expression microarry data. *Bioinformatics*, **2003**, *19*, 966−972.
(33) Box, G. E. P.; Cox, D. R. An Analysis of Transformations. *J. Royal Stat. Soc.* **1964**, *B-26*, 211−252.
(34) Cui X.; Churchill G. A. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **2003**, *4*, 210.
(35) Foster; A. M.; Tianm, L.; Wei, L. J. Estimation for the Box-Cox transformation model without assuming parametric error distribution, *J. Am. Stat. Assoc.* **2001**, *96*, 1097−1101.
(36) Geller, S. C.; Gregg, J. P.; Hagerman, P.; Rocke, D. M. Transformations and normalization of oligonucleotide microarray data. *Bioinformatics* **2003**, *19*, 1817−1823.
(37) Huber, W.; von Heydebreck, A.; Sultmann, H.; Poustka, A.; Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **2002**, *18*, 96−104.
(38) Cui, X.; Churchill, G. A. Transformations for cDNA microarray data. *Stat. App. Gen. Mol. Biol.* **2003**, *2*, 1,
(39) Draper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, 1982.
(40) McCulliagh, P.; Nelder, J. A. *Generalized Linear Models*; Chapman & Hall/CRC Press: Boca Raton, FL 1998
(41) Lawless, J. F. *Statistical Models and Methods for Lifetime Data*; Wiley Series in Probability and Statistics 2nd ed.; John Wiley & Sons: New York, 2003.
(42) Workshop Report: *Defining the Mandate of Proteomics in the Post-Genomics Era*; National Research Council: The National Academies Press: Washington, D. C., 2002.

PR0501261

International Conference on Diet, Nutrition, and Cancer

Chemoprevention of Breast Cancer, Proteomic Discovery of Genistein Action in the Rat Mammary Gland[1–3]

Craig Rowell,* D. Mark Carpenter,** and Coral A. Lamartiniere*†[4]

*Department of Pharmacology and Toxicology, and †UAB Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL 35294; and **Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849

ABSTRACT Genistein, the primary isoflavone component of soy, consumed in the diet during the prepubertal period only, and the combined prepubertal and adult periods, suppresses chemically induced mammary cancer in rats. Gestational or adult-only exposures do not provide protection. An inverse relation exists between cancer susceptibility and mammary gland differentiation. The current study used proteomic technology to investigate genistein mechanisms of action as related to programming against chemically induced mammary cancer. Rats were injected subcutaneously with 500 μg genistein/g body weight on d 16, 18, and 20 postpartum. At d 21, mammary glands were subjected to 2-dimensional polyacrylamide gel electrophoresis. After gel scanning, image analysis, and MS, 6 proteins were determined to be differentially regulated and identified. One protein, GTP-cyclohydrolase 1 (GTP-CH1), was confirmed as being significantly upregulated at d 21 by immunoblot analysis. Investigation of downstream signaling from GTP-CH1 showed that tyrosine hydroxylase was upregulated and vascular endothelial growth factor receptor 2 (VEGFR2) was downregulated in the mammary glands of 50-d-old rats treated with genistein in the prepubertal period. This and previous work suggest that early prepubertal exposure to genistein enhances cell proliferation by upregulating GTP-CH1 and the epidermal growth factor (EGF)-signaling pathway, and hence cell differentiation and gland maturation. This unique developmental maturation leads to a new biochemical blueprint, whereby the cells have reduced EGF signaling and VEGFR2, which renders the mature mammary gland less proliferative and less susceptible to cancer. This study demonstrated the usefulness of proteomics for the discovery of novel pathways that may be involved in cancer prevention. J. Nutr. 135: 2953S–2959S, 2005.

KEY WORDS: • *2-D gel* • *genistein* • *GTP-CH1* • *VEGFR2* • *proteomics*

Inherited genetic risk for breast cancer accounts for <10% of all breast cancer cases (1). Epidemiological reports and data from laboratory studies suggest that approximately three-fourths of all cancer deaths are attributable to lifestyle factors, indicating that environment and diet must play a role in the cause and prevention of cancer. Many cancer causation and prevention studies involving environmental chemicals have investigated direct effects in adult animals, because cancer is considered a disease associated with aging. However, exposure to environmental chemicals during critical periods of early development plays an important role in breast cancer susceptibility in adulthood (2).

Our laboratory has focused research on the chemopreventative nature of the soy isoflavone genistein. We demonstrated that prepubertal exposure to genistein decreases tumor multiplicity and diminishes incidence of adenocarcinomas in the dimethylbenz[a]anthracene (DMBA)[5] model of mammary cancer. Whole-mount analysis of mammary glands showed that prepubertal exposure to genistein enhances mammary gland differentiation, that is, results in fewer terminal end buds and more lobules (3,4). Terminal end buds are the least mature

[1] Published in a supplement to *The Journal of Nutrition*. Presented as part of the International Research Conference on Food, Nutrition, and Cancer held in Washington, DC, July 14–15, 2005. This conference was organized by the American Institute for Cancer Research and the World Cancer Research Fund International and sponsored by (in alphabetical order) California Avocado Commission; California Walnut Commission; Campbell Soup Company; The Cranberry Institute; Danisco USA, Inc.; The Hormel Institute; National Fisheries Institute; The Solae Company; and United Soybean Board. Guest editors for this symposium were Vay Liang W. Go, Ritva R. Butrum, and Helen A. Norman. *Guest Editor Disclosure*: R. R. Butrum and H. Norman are employed by conference sponsor American Institute for Cancer Research; V.L.W. Go, no relationships to disclose.

[2] Author disclosure: No relationships to disclose.

[3] The research was supported by grants from the U.S. Department of Defense (DOD BC 17-03-1-0433) and the National Institutes of Health (NIH U01 ES 012771-02). CR was supported by stipends from the U.S. Department of Defense (DOD BC 17-00-1-0119) and the Susan G. Komen Foundation (DISS0201242). The mass spectrometers used to generate the proteomic data were purchased using funds provided by Shared Instrumentation Grants (S10 RR11329 and S10 RR13795) from the National Center for Research Resources and an award from the UAB Health Services Foundation General Endowment Fund. The UAB Comprehensive Cancer Center Mass Spectrometry Shared Facility is supported by a core grant (P30 CA-13148) from the National Cancer Institute.

[4] To whom correspondence should be addressed. E-mail: Coral@uab.edu.

[5] Abbreviations: 2-D, 2-dimensional; BH4, tetrahydrobiopterin; BW, body weight; DMBA, dimethylbenz[a]anthracene; EB, estradiol benzoate; EGF, epidermal growth factor; GTP-CH1, GTP-cyclohydrolase 1; IHC, immunohistochemisty; MALDI-TOF, matrix-assisted laser desorption ionization–time of flight; MAPK, mitogen-activated protein kinase; VEGFR2, vascular endothelial growth factor receptor 2.

terminal ductal structures and are the most susceptible to chemical carcinogens (5). The resulting morphological action of genistein can be characterized as enhancing gland maturation. This maturation yields a gland that is ultimately less proliferative and less susceptible to chemical carcinogenesis (3,4,6,7). In general, the morphological development of the mammary gland extends beyond the embryonic stage and largely occurs during postnatal development. The postnatal development of the rat mammary gland closely resembles that of the human and provides a meaningful model for studying cancer chemoprevention (8).

To enhance our ability to discover biochemical pathways that genistein may affect in vivo, we adopted the technologies of 2-dimensional (2-D) gel electrophoresis and MS to find new biomarkers of genistein action in the rat mammary gland. In the present study, we discovered, via 2-D gel analysis, differential protein expression of GTP-cyclohydrolase I (GTP-CH1), the rate-limiting enzyme for the synthesis of tetrahydrobiopterin (BH4), a necessary cofactor for catecholamine synthesis and nitric oxide production (9). We also constructed a possible mechanism of chemoprevention by changes in the downstream enzyme tyrosine hydroxylase (responsible for dopamine production) and a possible target of its action, vascular endothelial growth factor receptor 2 (VEGFR2), and showed how these changes in protein expression can be involved in the chemopreventative action of genistein.

Methods and materials

Animals. Animal care and treatments were performed according to established guidelines, and protocols were approved by the UAB Animal Care Committee. Animals were housed in a temperature-controlled facility with a 12-h light:dark cycle. Female Sprague-Dawley rats (Charles River) were fed phytoestrogen-free AIN-76A pellets (Harlan Teklad) (10). Breeding was conducted in our facility, and animals were separated as soon as sperm or plugs were detected. Pregnant animals were caged separately. Offspring were collected and sexed on d 1 postpartum, and all litters were adjusted to 10 offspring (5 males and 5 females). Offspring were weaned at d 21 postpartum. Female offspring were assigned to the following groups: vehicle (dimethylsulfoxide; Sigma), genistein (98.5% purity; Hoffmann-LaRoche), estradiol benzoate (EB; Sigma), or daidzein (Hoffmann-LaRoche). On postnatal d 16, 18, and 20, female offspring were injected subcutaneously with either 500 μg genistein/g body weight (BW), 500 μg daidzein/g BW, 500 ng EB/g BW, or an equivalent volume of the vehicle. The genistein dose was previously shown to enhance mammary gland maturation and suppress chemically induced mammary cancer development (3). The daidzein dose was the same as the genistein dose, and the EB dose results in a similar uterotrophic effect as the dose of genistein (2,4,11). Animals were killed on d 21 or 50 postpartum (prepubertal and young adult rats, respectively). At the time of death, the 2 abdominal mammary glands were excised and snap-frozen at −80°C for later analysis. Three unique sets of genistein- and control-treated animals were generated for these experiments. Set 1 was used for the production of 2-D gels (5 rats/group); set 2 was used for validation studies via immunoblot analysis (8–10 rats/group). Set 3 included the daidzein and EB treatment groups (8 rats/group) as the weakly estrogenic isoflavone and estrogen-positive controls, respectively.

2-D gels. Frozen mammary tissues (n = 5, genistein; n = 5, control) were pulverized and homogenized in lysis buffer as described by Fountaoulakis et al. (12). Protein concentrations of the individual samples were determined using the BioRad protein assay reagent (BioRad) in 96-well microtiter plates. From each sample, 150 μg total protein was loaded onto 11-cm immobilized pH gradient strips, pH 3 to 10 (GE Healthcare). The samples were allowed to rehydrate into the gels overnight at room temperature. The next day, strips were placed onto a flatbed electrophoresis unit (Multiphor II; GE Healthcare) for 1st-dimension focusing. The voltage gradient was 500 V for 1 h, followed by a ramp to 3500 V for 1.5 h, and sustained at 3500 V for 22 h. After the 1st dimension was completed, the strips were equilibrated using 100 mmol/L dithiothreitol for 30 min followed by 30 min equilibration with 120 mmol/L iodacetimide. After equilibration, isoelectric focusing gels were placed on a 1.5-mm, 12.5% SDS vertical gel (Criterion; BioRad), and the gels were assayed on a vertical electrophoresis unit (Dodecacell, BioRad) at a constant 200 V. When the assays were completed, the gels were fixed for 1 h in 40% methanol and 10% acetic acid (v:v), and stained overnight with Sypro Ruby gel stain (Molecular Probes). Gels were destained (10% methanol, 7.5% acetic acid; v:v) for a minimum of 4 h. Stained gels were scanned via a Perkin Elmer ProExpress densitometer and initially analyzed with the Progenesis 2-D gel software system (Nonlinear).

MS. For identification, selected protein spots were manually excised from the gels. Gel plugs were destained, dried in a SpeedVac (Savant), rehydrated, and digested with trypsin (Roche). Samples were mixed 1:10 (v:v) with a saturated solution of sinapinic acid in 50% aqueous acetonitrile and 0.1% aqueous trifluoroacetic acid (1:1, v:v), and 1 μL was spotted onto the stainless steel matrix-assisted laser desorption ionization (MALDI) target plate and allowed to dry before analysis by MALDI–time-of-flight (TOF) MS. Peptide molecular ions were analyzed in linear positive-ion mode using a Voyager Elite MS (Applied Biosystems). Using an acceleration voltage of 25 kV and a laser intensity of 2500 V, each spot was analyzed a minimum of 3 times, accumulating spectra composed of ~200 laser shots in total. The resulting spectra were analyzed by DataExplorer (Applied Biosystems). The instrument was calibrated using an external apomyoglobin standard and internally using a trypsin autolysis peak. Resulting spectra were baseline corrected and filtered for noise. MALDI-TOF MS–produced spectra were analyzed using the MASCOT program (13).

Immunoblot analysis. Mammary gland samples (n = 8/treatment) were homogenized in radioimmunoprecipitation lysis buffer (Upstate Biotech) with protease inhibitors (2 mmol/L Na vanadate, 0.2 mmol/L phenylmethylsulfonyl fluoride, 2 μg leupeptin/mL, 2 μg aprotinin/mL) while rotating at 4°C. Proteins from each sample (20 μg) were separated by SDS-PAGE and transferred to a nitrocellulose membrane (BioRad). Membranes were blocked and incubated with the appropriate antibody, including polyclonal rabbit antirat GTP-CH1 (Dr. Gregory Kapotos, Wayne State University), and the commercially available tryosine hydroxylase, tryptophan hydroxylase, phenylalanine hydroxylase, inducible nitric oxide synthase, and VEGFR2 (Santa Cruz). Membranes were subsequently incubated with appropriate secondary antibody conjugated to horseradish peroxidase for detection after reaction with a chemiluminescent substrate (Pierce). Membranes were either exposed to autoradiography film for band quantitation by densitometric analysis or directly analyzed using a VersaDoc 4000 (BioRad). Densitometric analysis was performed using QuantityOne software (BioRad).

ELISA. Dopamine levels were evaluated using an ELISA kit (Rocky Mountain Diagnostics) according to the manufacturer's protocol. Briefly, whole mammary glands were homogenized in RIPA buffer with protease inhibitors; 10 μL from each sample was used per evaluation, and all samples were analyzed in duplicate in 96-well microtiter plates. The range of standards was from 0 to 2560 μg/L, with a total of 6 separate concentrations. Samples were evaluated using an Ultrospec 2000 UV/Visible spectrophotometer (GE Healthcare, formerly Pharmacia Biotech).

Immunohistochemisty (IHC). Immunolocalization of tyrosine hydroxylase, GTP-CH1, and VEGFR-2 was performed as described in Brown et al. (14) on trypsinized paraffin sections of mammary glands fixed in 4% paraformaldehyde by a modified avidin-biotin complex technique. Primary antibodies were the same as described for the immunoblot methods plus mouse monoclonal dopamine (Abcam). All primary antibodies were incubated overnight at 4°C in a humidity chamber with rocking. The secondary antibodies were biotinylated goat antirabbit IgG (Vector Laboratories). Negative control slides were treated with preimmune serum and omission of primary antibody with each batch of slides. Slides were evaluated on a Nikon Labphot-2 (Nikon) at 200× magnification (10× ocular and 20× objective).

HPLC. BH4 levels were assayed using HPLC methods described by Fukushima and Nixon (15). Briefly, whole mammary glands (n

= 10/treatment) were lysed in 0.15 mol/L $HClO_4$ and 0.1 mol/L H_3PO_4. An internal standard of D-erythro-neopterin (gift of Dr. Wayne Kapatos, Wayne State University) was added to each sample. To each sample 300 μL of a 1% iodine and 2% potassium iodide solution was added, and the samples were incubated in the dark for 1 h. A solution of 1% ascorbic acid was then added to stop the reaction. Proteins were precipitated, and the supernatant was passed through a Dowex 50W-X4 column (Bio-Rad). Pteridines were eluted from the column using 1 mol/L NH_4OH directly onto a Dowex AG1-X8 column (Bio-Rad). Finally, pteridines were eluted using 1 mol/L acetic acid, then dried by evaporation. The samples were analyzed on a Perkin Elmer Series 200 autosampler and pump with a Shimadzu RF-551 spectrofluorometric detector using an Alltech Associates Spherisorb ODS-1 4.6-mm-i.d. × 250-mm column. The mobile phase was 5% methanol in water at 1 mL/min.

Statistics. For 2-D gels, the average values for normalized spot volume were compared using the Student's t test. Average values of immunoblots for each treatment group were compared using ANOVA combined with Tukey's test (SAS) and reported as percentage of the control.

Results

Proteins from the mammary glands of 21-d-old rats with prepubertal exposure to genistein or control treatments were evaluated using 2-D gel electrophoresis. After image analysis, 88 proteins common to all gels were selected for further statistical evaluation. Based on normalized spot volume, the expression of 6 protein spots was found to differ significantly between the genistein-treated and the control animals ($P < 0.05$) (Fig. 1). Table 1 lists the protein names and their descriptive roles(16–20). For 5 of these proteins (B–F), we were able to obtain antibodies and pursue protein confirmation. Immunoblot analysis of the individual proteins confirmed only the change observed for GTP-CH1 (Fig. 2).

After identifying and confirming changes in GTP-CH1 expression, we used extensive literature and Web searching [e.g., Tetrahydrobiopterin Home Page (21)] to find information for related pathways that might explain how this particular protein might play a role in mammary cancer chemoprevention. We found that GTP-CH1 is the rate-limiting step in

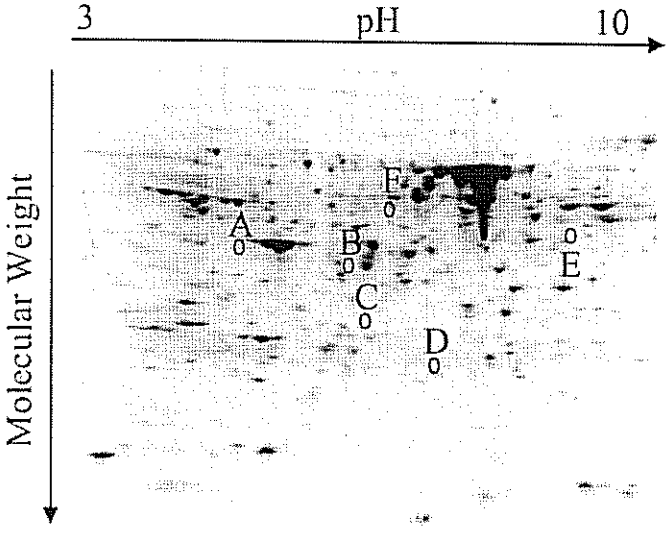

FIGURE 1 Representative 2-D gel showing the location of protein spots that differed significantly in expression based on normalized spot volume in the mammary glands of 21-d-old rats with prepubertal genistein treatment: A, fertility protein SP22; B, γ-synuclein; C, ABC transporter; D, peroxiredoxin 1; E, 14–3-3 ε; F, GTP-CH1.

the production of BH4, which is a necessary cofactor for the activity of the amino acid hydroxylases (tryosine hydroxylase, tryptophan hydroxylase, and phenylalanine hydroxylase) and the nitric oxide synthases. These enzymes play a key role in the production of catecholamines as well as nitric oxide (9).

BH4 levels were measured by HPLC (Fig. 3). Although mean values for BH4 from the mammary glands of treated and control animals did not differ significantly, the overall variance in expression as evaluated using the Weibel model was significantly different ($P < 0.05$).

Protein levels of the amino acid hydroxylases in the mammary glands of 21-d-old rats treated with genistein were evaluated by immunoblot analysis (Fig. 4). None of the examined proteins, except GTP-CH1, differed significantly at this time. To determine whether the changes in GTP-CH1 were more than a direct effect of genistein action, we evaluated all proteins at 50 d. Results of immunoblots from mammary glands showed a significant increase only in levels of tyrosine hydroxylase ($P < 0.05$) (Fig. 5).

After determining that tyrosine hydroxylase levels were increased, we measured one of its products, dopamine, and a protein known to be regulated by dopamine, VEGFR2. Dopamine levels did not differ between the genistein and control groups (data not shown). VEGFR2 was downregulated in the mammary glands of 50-d-old rats (Fig. 6).

Given the importance of the microenvironment in signaling and development, we used IHC to provide information about the distribution of the 3 proteins that were differentially regulated. Results of IHC for the proteins GTP-CH1, tyrosine hydroxylase, and VEGFR2 showed that these biomolecules were confined to the epithelial cells of the mammary structures (Fig. 7). Staining intensity confirmed immunoblotting analysis for these proteins.

Given the structural activity relation of genistein with estrogen, we evaluated whether the observed changes in protein expressions were likely due to the estrogenic action of genistein. Animals were treated with 500 μg daidzein/g BW or 500 ng EB/g BW. Daidzein is a soy isoflavone not associated with chemoprevention (22). EB at the given concentration causes a uterotrophic effect in rats. Results from this study showed that prepubertal exposure to daidzein or EB did not significantly alter GTP-CH1, tyrosine hydroxylase, and VEGFR2 levels as measured by immunoblot analysis in mammary glands of 21- and 50-d-old rats (data not shown).

Discussion

We previously reported that short-term treatment of rats during the prepubertal period with genistein can confer a long-term protection against chemically induced mammary cancer (2–4,6). The cellular mechanism of action was enhanced mammary gland maturation. Until recently, few biochemical mechanisms have been associated with this mammary cancer chemoprevention model (14). Hence, we embarked on the use of 2-D gels for finding additional differentially expressed proteins and MALDI-TOF MS for identifying these proteins.

Our first endeavor into discovery proteomics revealed 6 protein spots from 2-D gels that were differentially expressed in the mammary glands of 21-d-old rats treated with genistein in the prepubertal period. However, only protein expression change for GTP-CH1 was confirmed by immunoblot analysis. GTP-CH1 expression was upregulated in the mammary glands of 21-d-old but not in 50-d-old rats with only prepubertal exposure to genistein. This demonstrates a direct and reversible action, because GTP-CH1 levels were unchanged in the mammary glands of the resulting adult rats.

TABLE 1

Proteins found to be differentially expressed in mammary glands of rats with prepubertal genistein exposure by 2-D gel electrophoresis[1]

Spot	Protein name, accession number	Mowse score	Differential expression	Putative activity
A	Fertility protein SP22 AAD43957	80	Upregulated	Protein predictive of fertility (16)
B	γ-Synuclein NP_113876 XP_346482	91	Upregulated	Enhances binding affinity of the estrogen receptor (17)
C	ABC transporter CAA05793	83	Upregulated	Implicated in MDR (18)
D	Peroxiredoxin 1 NP_476455	94	Downregulated	Antioxidant enzyme, may act as tumor suppressor in older animals (19)
E	14-3-3 ε NP_113791	81	Upregulated	Tyrosine hydroxylase activator (20)
F	GTP-CH1 PC22 56	105	Upregulated	Rate-limiting step in production of BH4 (9)

[1] Protein identification from peptide mass-fingerprinting results of MALDI-TOF MS, analyzed using the Mascot search engine.

To determine the biochemical significance of GTP-CH1 upregulation in the immature mammary gland, we investigated downstream metabolic pathways (**Fig. 8**). GTP-CH1 is the rate-limiting enzyme in the production of BH4. The latter is an essential cofactor for the enzymes of catecholamine and nitric oxide syntheses. Catecholamines, in particular, can signal through membrane-bound adrenergic receptors to affect differentiation and development of cells (23). In PC12 cells, stimulation by epidermal growth factor or nerve growth factor increases cell proliferation through elevation of intracellular BH4 (24). BH4 is also implicated in cell proliferation on the basis of studies using murine erythroleukemia cells as a model for erythrogenesis (25). Although we did not find a direct correlation between GTP-CH1 levels and BH4 levels, we did notice a difference in the overall regulation of BH4 between the 2 groups, based on variance. This is not conclusive evidence of the actions of GTP-CH1, but others, such as Vann et al. (26), have reported a divergence in the regulation of BH4 and enzymes that require it (primarily nitric oxide synthase in their experiment).

Consistent with the reported action of BH4 on cell proliferation and differentiation (23,24), our earlier reports showed that genistein administration to prepubertal rats results in an initial increase in the EGF-signaling pathway and cell proliferation in the mammary glands of 21-d-old rats, followed by increased mammary gland maturation and then decreased EGF signaling in mammary the glands of 50-d-old rats with only prepubertal exposure to genistein (2,3,14). More recently, we found that phosphorylated extracellular regulated kinase-1 was significantly upregulated in the mammary glands of 21- but not 50-d-old rats (unpublished data, 2005). Hence, catecholamine action may enhance genistein's possible direct action on EGF-signaling and mitogen-activated protein kinase (MAPK) signaling to temporally increase cell proliferation in prepubertal rats and enhance cell differentiation. A review by Cobb (27) highlights some of the multiple signaling mechanisms that affect MAPK signaling and are likely involved in our observed morphological changes. Earlier studies by Matsuda et al. (28) determined a role for catecholamines in the mammary gland during the first stages of involution.

Given the significant increase in GTP-CH1 expression and the potential involvement of its downstream targets, we investigated whether the direct effects of genistein altered the expression of the catecholamine synthesis enzymes. Immunoblot analysis showed that only the level of tyrosine hydroxylase was significantly altered. Specifically, we found that it was upregulated

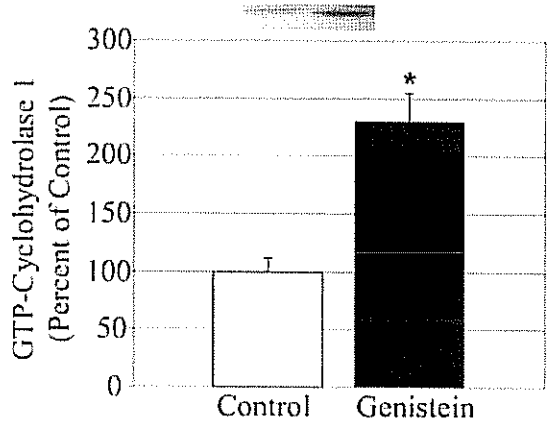

FIGURE 2 Immunoblot analysis for GTP-CH1 in mammary glands of 21-d-old rats with and without prepubertal genistein: immunoblots (*top*) and graph of densitometry measurements from these immunoblots. Densitometry values for controls were set to 100. Values are means ± SEM (n = 8). *Different from controls, $P < 0.05$.

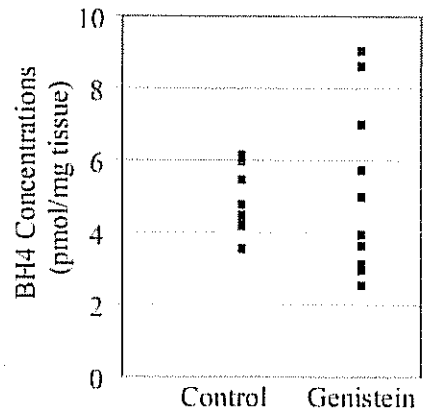

FIGURE 3 BH4 levels from mammary glands of 21-d-old rats with prepubertal genistein exposure (n = 10/group).

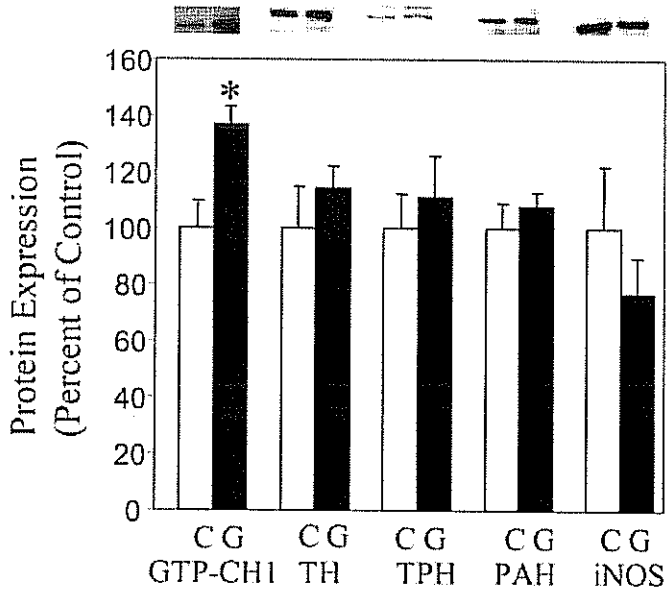

FIGURE 4 Immunoblot analysis for GTP-CH1, tyrosine hydroxylase (TH), tryptophan hydroxylase (TPH), phenylalanine hydroxylase (PAH), and inducible nitric oxide synthase (iNOS) in mammary glands of 21-d-old rats with (G) and without (C) prepubertal genistein treatment: immunoblots (*top*) and graph of densitometry measurements from these immunoblots. Densitometry values for controls were set to 100. Values are means ± SEM (n = 8). *Different from controls, $P < 0.05$.

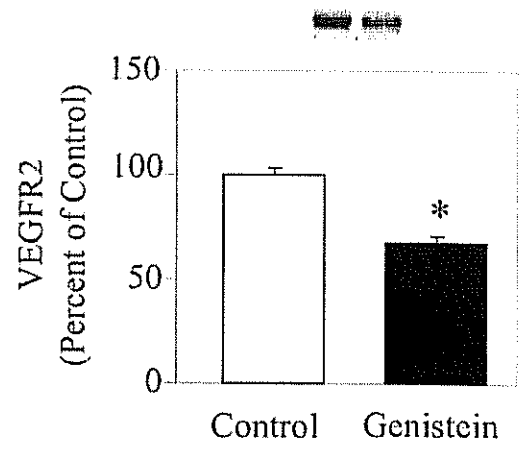

FIGURE 6 Immunoblot analysis for VEGFR2 in mammary glands of 50-d-old rats with prepubertal genistein treatment: immunoblots (*top*) and graph of densitometry measurements from these immunoblots. Densitometry values for controls were set to 100. Values are means ± SEM (n = 8). *Different from controls, $P < 0.05$.

in mammary glands of 50- but not 21-d-old rats. Notably, this change in protein level at 50 d occurred in the absence of the original effector, genistein, and was present even though the significant difference in the level of GTP-CH1 no longer existed. This is suggestive of a programming effect on protein expression,

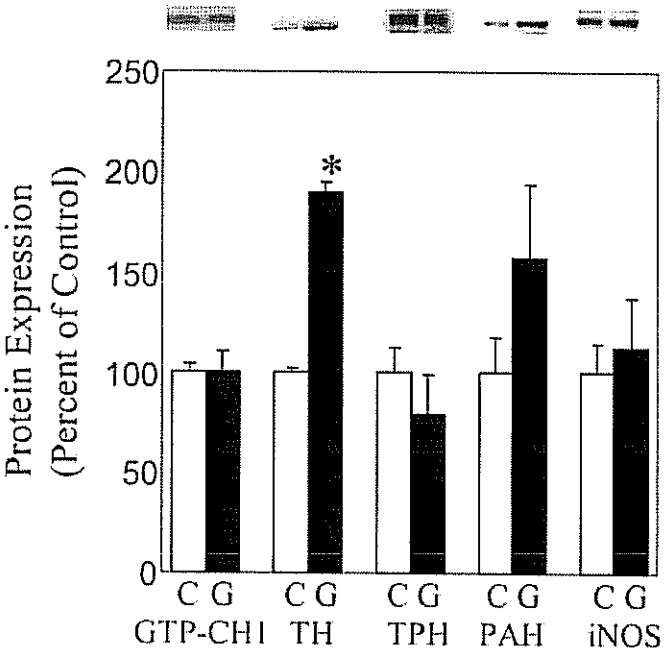

FIGURE 5 Immunoblot analysis for GTP-CH1, tyrosine hydroxylase (TH), tryptophan hydroxylase (TPH), phenylalanine hydroxylase (PAH), and inducible nitric oxide synthase (iNOS) in mammary glands of 50-d-old rats with (G) and without (C) prepubertal genistein treatment: immunoblots (*top*) and graph of densitometry measurements from these immunoblots. Densitometry values for controls were set to 100. Values are means ± SEM (n = 8). *Different from controls, $P < 0.05$.

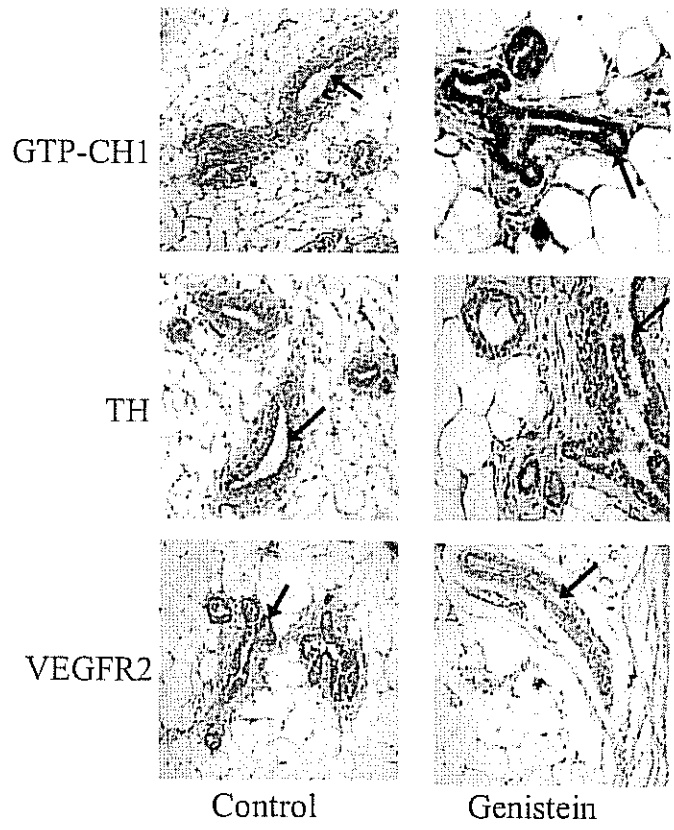

FIGURE 7 Representative immunohistochemistry for GTP-CH1 in mammary glands of 21-d-old rats and tyrosine hydroxylase (TH) and VEGFR2 in mammary glands of 50-d-old rats. Note brown immunostaining for GTP-CH1, TH, and VEGFR2. These proteins were localized to the epithelium (*arrows*).

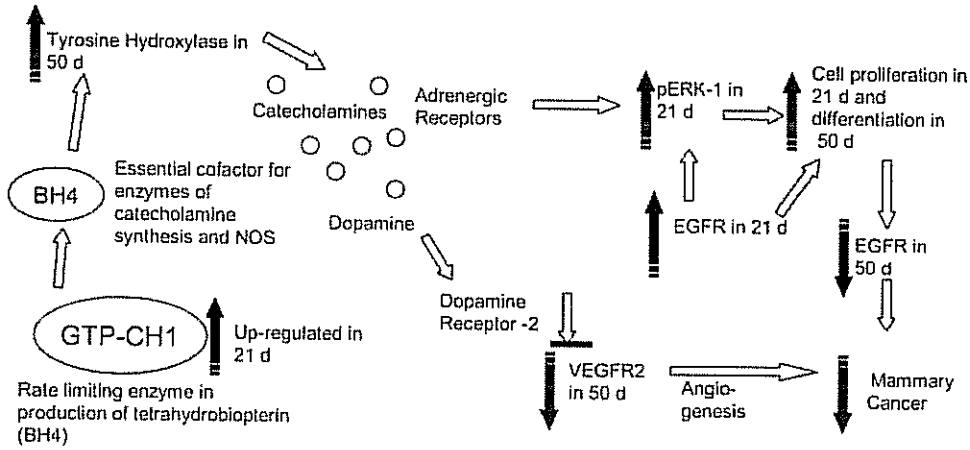

FIGURE 8 Proposed mechanism of action of genistein in regulating mammary gland maturation, cell differentiation, and mammary cancer susceptibility. Open arrows indicate pathways; solid arrows indicate biological action measured in mammary glands of rats with prepubertal genistein exposure.

that is, a developmental modification with no observable effect on the protein immediately after treatment or exposure but a delayed manifestation in adult life (2). In light of this change, we turned our attention to how downstream metabolic pathways and signaling of tyrosine hydroxylase in the mature gland could be involved in suppressing cancer in rats with only prepubertal exposure to genistein.

Teunis et al. (29) examined the role of dopamine sensitivity in tumor growth. They found that rats with high dopaminergic activity had a reduction in tumor size compared with rats with low dopaminergic reactivity. Associating elevated dopamine levels with suppressed mammary tumorigenesis, they noted that the angiogenic response to VEGF could be inhibited by administration of dopaminergic agonists. Basu et al. (30) reported that dopamine acts through the dopamine 2 receptor to induce the endocytosis of VEGFR2 and thereby inhibit or prevent VEGF binding, receptor phosphorylation, and subsequent signaling steps. They reported that immunohistochemical studies did not find tyrosine hydroxylase–positive nerves in tumors, and the dopamine concentration in malignant tumors was significantly reduced compared with concentrations in controls. Ferguson et al. (31) reported that lifetime exposure to genistein (500 μg/g) can potentiate dopamine levels in striata of amphetamine-exposed animals. This demonstrates a lasting effect of genistein that does not affect the baseline levels but allows for adjustments in the face of certain perturbations. In addition, researchers found that genistein decreased both transcription and protein levels of VEGF and that this decrease is involved in the loss of angiogenesis (32,33). Hefelfinger et al. (34) demonstrated that inhibition of VEGFR2 will prevent DMBA-induced mammary tumors in rats.

Given a potential link among tyrosine hydroxylase, VEGFR2, and angiogenesis, we conducted analysis for the VEGFR2 protein. The results showed a significant decrease in VEGFR2 expression in the mammary glands of 50-d-old rats with only prepubertal genistein treatment. These results complemented the finding that tyrosine hydroxylase levels were elevated in the mammary glands of 50-d-old rats with prepubertal genistein treatment. We speculated that upregulation of tyrosine hydroxylase results in dynamic upregulation of catecholamines, which, in turn, decrease the VEGFR2 levels, resulting in decreased ability to promote angiogenesis (Fig. 8). Ortega et al. (35) implicated VEGFR2 in mediating cell proliferation; therefore, a decrease in VEGFR2 may decrease the overall proliferative potential of the mammary gland. The absence of a demonstrable change in dopamine concentrations may mean that the concentration is dynamic or that changes in concentration within the microenvironment may not manifest in detectible or significant change in the larger sample (whole mammary gland).

Having established possible mechanisms by which the direct effects of genistein on GTP-CH1 and the programming effects of genistein on tyrosine hydroxylase could lead to developmental changes in the mammary glands and alter susceptibility to breast cancer, we deemed it important to determine whether the observed effects resulted from actions unique to genistein or were part of a broader estrogenic response. To answer this question, we investigated how selected proteins respond in animals exposed to estrogen or the other major soy isoflavone, daidzein. Results from the present study showed that prepubertal exposure to EB and daidzein did not produce any significant changes for GTP-CH1, tyrosine hydroxylase, VEGFR2, or any of the other evaluated proteins in the mammary glands of 21- and 50-d-old rats. This suggested that genistein exhibits some selective effect independent of estrogen action. The lack of significant changes brought about by daidzein as compared with genistein supports the role of these changes as unique potential mechanisms for genistein chemoprevention.

Localization of the proteins of interest to the same cell type may provide greater insight into the role of these proteins within this model. Our IHC data indicate that the proteins of interest (GTP-CH1, tyrosine hydroxylase, and VEGFR2) appear to be modulated in their expression only in the lining of epithelial cells. This is important because it is hypothesized that the epithelial cell type may be the origination of tumor cells. Because all 3 proteins are expressed in the same cell type and IHC confirmed that these changes were similar to those observed for immunoblot analysis, this may indicate a paracrine or an autocrine regulatory action. This colocalization will deserve further examination as we explore the temporal expression of these proteins in this model.

Although we were able to find several proteins determined to be differentially regulated using 2-D gel analysis, our confirmation attempts using immunoblots found that only one of these proteins was significantly changed by genistein treatment. This lack of cohesion between our 2-D gels and immunoblotting analysis may be because of stringent specificity of site-directed antibodies and/or due to the low number of samples used for the generation of the 2-D gel data set. We used $n = 5$ for the 2-D gels, but our confirmation was with $n = 8$/group. Given that the sensitivity of the immunotechniques should be much greater than that of the staining techniques, we recognize that we must increase the number of samples in subsequent 2-D gel studies to decrease this false-

positive rate. However, we recognize that a finite number of samples can be reasonably assayed to produce quality data without exhausting resources. Also, the use of *t*-tests and other parametric procedures may not be appropriate mechanisms for evaluating the 2-D gel data set; therefore, alternative testing strategies may be needed for evaluating potential subtle changes in expression. These evaluations may focus on general changes of protein expression rather than specific changes related to mean expression. For these reasons, we are still interested in pursuing the other identified proteins.

In summary, data from this and previous work suggest that early postnatal (prepubertal) exposure to genistein enhances cell proliferation by upregulating GTP-CH1 and the EGF-signaling pathway and hence cell differentiation and gland maturation. This unique developmental maturation leads to a new biochemical blueprint, whereby the cells have reduced EGF-signaling and VEGFR2, which renders the mature mammary gland less proliferative and less susceptible to chemically induced mammary cancer initiation, angiogenesis, and cancer progression. Therefore, genistein acts through a diverse and coordinated effect of signaling mechanisms and pathways that likely account for the cellular changes responsible for its chemopreventive action.

Because cancer is often thought of as a disease of aging, we have to understand and appreciate that the mammary gland undergoes multiple stages of development. The use of animal models allows us to follow the changes within the gland to determine how the actions of genistein may alter the morphological characteristics as well as to examine how these morphological changes may work at the biochemical level to yield sustained changes that affect susceptibility. Two-dimensional gel exploration looks at the most common and persistent changes within the cellular populations. Knowledge gained from these experiments provides more information on the possible mechanisms of action of genistein that result in chemoprevention. Specifically, this report is the first to identify GTP-CH1 protein alterations in the mammary gland as well as their potential regulation by genistein. Future work will examine other members of these signaling pathways more closely and address the potential mechanisms involved.

ACKNOWLEDGMENTS

We thank Gregory Kapatos at Wayne State University, Detroit, MI, for providing the GTP-CH1 antibody, and Stephen Barnes and Landon Wilson at the UAB Comprehensive Cancer Center Mass Spectrometry Core Facility for their assistance with MS protein identification. We thank Isam-Eldin Eltoum for his assistance with the histomorphology and Jun Wang for general technical assistance. As well, we thank Lori Coward at Southern Research, Birmingham, AL, for measurement of BH4 levels.

LITERATURE CITED

1. Colditz GA, Frazier AL. Models of cancer show that risk is set by events of early life: prevention efforts must shift focus. Cancer Epidemiol Biomarkers Prev. 1995;4:567–71.

2. Lamartiniere CA. Timing of exposure and mammary cancer risk. J Mammary Gland Biol Neoplasia. 2002;7:67–76.

3. Murrill WB, Brown NM, Zhang J-X, Manzolillo PM, Barnes S, Lamartiniere CA. Prepubertal genistein exposure suppresses mammary cancer and enhances gland differentiation in rats. Carcinogenesis. 1996;17:1451–7.

4. Fritz WA, Coward L, Wang J, Lamartiniere CA. Dietary genistein: perinatal mammary cancer prevention, bioavailability and toxicity testing in the rat. Carcinogenesis. 1998;19:2151–8.

5. Russo J, Wilgus G, Russo IH. Susceptibility of the mammary gland to carcinogenesis. I. Differentiation of the mammary gland as determinant of tumor incidence and type of lesion. Am J Pathol. 1979;96:721–34.

6. Lamartiniere CA, Moore JB, Brown NM, Thompson R, Hardin MJ, Barnes S. Genistein suppresses mammary cancer in rats. Carcinogenesis. 1995;16: 2833–40.

7. Lamartiniere CA, Murrill WB, Manzolillo PA, Zhang J-X, Barnes S, Zhang X, Wei H, Brown NM. Genistein alters the ontogeny of mammary gland development and protects against mammary cancer in rats. Proc Soc Exper Biol Med. 1998;217:358–64.

8. Russo J, Russo IH. Toward a physiological approach to breast cancer prevention. Cancer Epidemiol Biomarkers Prev. 1994;3:353–64.

9. Nomura T, Tazawa M, Ohtsuki M, Sumi-Ichinose C, Hagino Y, Ota A, Nakashima A, Mori K, Sugimoto T, et al. Enzymes related to catecholamine biosynthesis in *Tetrahymena pyriformis*. Presence of GTP cyclohydrolase I. Comp Biochem Physiol B Biochem Mol Biol. 1998;120:753–60.

10. Bieri JG, Stoewsand GS, Briggs GM, Phillips RW, Woodward JC, Knapka JJ. Report of the American Institute of Nutrition Ad Hoc Committee on Standards for Nutritional Studies. J Nutr. 1980;110:1340–8.

11. Cotroneo MS, Wang J, Eltoum IE, Lamartiniere CA. Sex steroid receptor regulation by genistein in the prepubertal rat uterus. Mol Cell Endocr. 2001;173: 135–45.

12. Fountaoulakis M, Schull E, Hardmeier R, Bernt P, Lubec G. Rat brain proteins: Two-dimensional protein database and variations in the expression level. Electrophoresis. 1999;20:3572–9.

13. MASCOT [program on the Internet]. Boston: Matrix Science Inc. c2003 [updated 2005 May 3; cited 2005 Aug 3]. Available from: http://www.matrixscience.com.

14. Brown NM, Wang J, Cotroneo MS, Zhao YX, Lamartiniere CA. Prepubertal genistein treatment modulates TGF-alpha, EGF and EGF-receptor mRNAs and proteins in the rat mammary gland. Mol Cell Endocrinol. 1998;144:149–65.

15. Fukushima T, Nixon JC. Chromatographic analysis of pteridines. In: McCormick DB, Wright LD, editors. Methods in enzymology, 66th ed. San Diego (CA): Academic Press; 1980. p. 429–36.

16. Klinefelter GR, Laskey JW, Ferrell J, Suarez JD, Roberts NL. Discriminant analysis indicates a single sperm protein (SP22) is predictive of fertility following exposure to epididymal toxicants. J Androl. 1997;18:139–50.

17. Jiang Y, Liu YE, Goldberg ID, Shi YE. Gamma synuclein, a novel heat-shock protein-associated chaperone, stimulates ligand-dependent estrogen receptor alpha signaling and mammary tumorigenesis. Cancer Res. 2004;64:4539–46.

18. Leonessa F, Clarke R. ATP binding cassette transporters and drug resistance in breast cancer. Endocr Relat Cancer. 2003;10:43–73.

19. Noh DY, Ahn SJ, Lee RA, Kim SW, Park IA, Chae, HZ. Overexpression of peroxiredoxin in human breast cancer. Anticancer Res. 2001;21:2085–90.

20. Hermeking H. The 14–3-3 cancer conection. Nat Rev Cancer. 2003;3: 931–43.

21. Tetrahydrobiopterin Home Page [homepage on the Internet]. Jona, Switzerland: Biomarin, Schircks Laboratories [cited 2005 Aug 19]. Available from: http://www.bh4.org/.

22. Lamartiniere CA, Wang J, Smith-Johnson M, Eltoum IE. Daidzein: bioavailability, potential for reproductive toxicity, and breast cancer chemoprevention in female rats. Toxicol Sci. 2002;65:228–38.

23. Schallreuter KU, Lemke KR, Pittelkow M, Wood J, Korner C, Malik R. Catecholamines in human keratinocyte differentiation. J Invest Dermatol. 1995; 104:953–7.

24. Anastasiadis PZ, Bezin L, Imerman BA, Kuhn DM, Louie MC, Levine RA. Tetrahydrobiopterin as a mediator of PC12 cell proliferation induced by EGF and NGF. Eur J Neurosci. 1997;9:1831–7.

25. Tanaka K, Kaufman S, Milstien S. Tetrahydrobiopterin, the cofactor for aromatic amino acid hydroxylases, is synthesized by and regulates proliferation of erythroid cells. Proc Natl Acad Sci U S A. 1989;86:5864–7.

26. Vann LR, Twitty S, Spiegel S, Milstien S. Divergence in regulation of nitric-oxide synthase and its cofactor tetrahydrobiopterin by tumor necrosis factor-alpha. Ceramide potentiates nitric oxide synthesis without affecting GTP cyclohydrolase I activity. J Biol Chem. 2000;275:13275–81.

27. Cobb MH. MAP kinase pathways. Prog Biophys Mol Biol. 1999;71:479–500.

28. Matsuda M, Imaoka T, Vomachka AJ, Gudelsky GA, Hou Z, Mistry M, Bailey JP, Nieport KM, Walther DJ, et al. Serotonin regulates mammary gland development via an autocrine-paracrine loop. Dev Cell. 2004;6:193–203.

29. Teunis MA, Kavelaars A, Voest E, Bakker JM, Ellenbroek BA, Cools AR, Heijnen CJ. Reduced tumor growth, experimental metastasis formation, and angiogenesis in rats with a hyperreactive dopaminergic system. FASEB J. 2002; 16:1465–7.

30. Basu S, Nagy JA, Pal S, Vasile E, Eckelhoefer IA, Bliss VS, Manseau EJ, Dasgupta PS, Dvorak HF, Mukhopadhyay D. The neurotransmitter dopamine inhibits angiogenesis induced by vascular permeability factor/vascular endothelial growth factor. Nat Med. 2001;7:569–74.

31. Ferguson, SA, Flynn KM, Delclos KB, Newbold RR, Gough, BJ. Effects of lifelong dietary exposure to genistein or nonylphenol on amphetamine-stimulated striatal dopamine release in male and female rats. Neurotoxicol Teratol. 2002;24: 37–45.

32. Fotsis T, Pepper M, Adlercreutz H, Hase T, Montesano R, Schweigerer L. Genistein, a dietary ingested isoflavonoid, inhibits cell proliferation and in vitro angiogenesis. J Nutr. 1995;125:790S-7.

33. Mukhopadhyay D, Tsiokas L, Zhou XM, Foster D, Brugge JS, Sukhatme VP. Hypoxic induction of human vascular endothelial growth factor expression through c-Src activation. Nature. 1995;375:577–81.

34. Heffelfinger SC, Yan M, Gear RB, Schneider J, LaDow K, Warshawsky D. Inhibition of VEGFR2 prevents DMBA-induced mammary tumor formation. Lab Invest. 2004;84:989–98.

35. Ortega N, Hutchings H, Plouet J. Signal relays in the VEGF system. Front Biosci. 1999;4:D141–52.